

must be reproduced?

(7M)

Release 3.1A John F. Collins, Biocomputing Research Unit
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

```
Run on:      Wed Apr 19 23:38:47 2000;      MasPar time 17.05 Seconds
Tabular output not generated.      860.240 Million cell updates/sec
```

```

Title: >US-09-323-597-2
Description: (1491) from US09323597.pep
Perfect Score: 3673
Sequence: 1 MALNSSSPAIGYIYENHGV.....VGNMVFETDNIYQMRADG 491

```

Scoring table: PAM 150
Gap 11

Searched: 82229 seqs, 29864866 residues

```
Post-processing: Minimum Match 08
                  Listing first 45 summaries
```

Database: swiss-prot38
1:swissprot

Statistics: Mean 48.170; Variance 69.227; scale 0.696

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description	Pred. No.
1	3635	99.0	492	1	TMS2_HUMAN	0.00e+00
2	895	24.4	1035	1	ENTR_BOVIN	1.37e-21
3	890	24.2	1034	1	ENTR_HUMAN	3.70e-21
4	862	23.5	1013	1	ENTR_PIG	3.79e-20
5	845	23.0	1069	1	ENTR_MOUSE	2.73e-19
6	762	20.7	638	1	KAL_HUMAN	1.15e-17
7	740	20.1	638	1	KAL_RAT	1.97e-16
8	720	19.6	638	1	KAL_MOUSE	6.98e-16
9	694	18.9	625	1	FA1L_HUMAN	1.96e-15
10	660	18.5	417	1	HEPS_HUMAN	1.72e-15
11	660	18.0	416	1	HEPS_MOUSE	7.22e-14
12	658	17.9	902	1	EPIT_MOUSE	2.63e-14
13	652	17.8	416	1	HEPS_RAT	1.27e-14
14	634	17.3	812	1	PLMN_MOUSE	4.14e-13
15	617	16.8	338	1	PLMN_HORSE	8.01e-13
16	613	16.7	790	1	PLMN_PIG	1.05e-13
17	608	16.6	810	1	PLMN_MACQU	4.96e-13
18	607	16.5	270	1	TRYT_MERUN	2.96e-13
19	597	16.3	810	1	PLMN_HUMAN	3.04e-12
20	596	16.2	343	1	PLMN_SHEEP	5.78e-12
21	587	16.0	273	1	MCRT_RAT	1.85e-12
22	586	16.0	333	1	PLMN_CANFA	3.51e-12
23	583	15.9	267	1	TRYT_ANOGA	2.40e-12
					TRYPIN 7 PRECURSOR (E	

RESULT	1	ALIGNMENTS
ID	TMS2_HUMAN	STANDARD; PRT; 492 AA.
AC	015393;	
DT	15-JUL-1998 (Rel. 36, Created)	
DT	15-JUL-1998 (Rel. 36, Last sequence update)	
DT	15-JUL-1998 (Rel. 36, Last annotation update)	
DE	TRANSMEMBRANE PROTEASE, SERINE 2 (EC 3.4.21.-).	
GN	TPRSS2.	
OS	Homo sapiens (Human).	
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
CC	Eutheria; Primates; Catarrhini; Homiidae; Homo.	
RN	SEQUENCE FROM N.A.	
RA	PAOLINI-GIACOBINO A., CHEN H., PETTSCH M.C., ROSSIER C.,	
RA	ANTONARAKIS S.E.;	
RT	"Cloning of the TPRS2 gene, which encodes a novel serine protease	
RL	with transmembrane, LDLRA, and SRR domains and maps to 21q22.3.";	
CC	Genomics 44:309-320(1997).	
CC	-1 SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.	
CC	-1 TISSUE SPECIFICITY: EXPRESSED STRONGLY IN SMALL INTESTINE AND	
CC	WEAKLY IN SEVERAL OTHER TISSUES.	
CC	-1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE	
CC	TRYPsin FAMILY.	
CC	-1 SIMILARITY: CONTAINS 1 SRR DOMAIN.	
CC	-1 SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation	
CC	at the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (see http://www.isb-sib.ch/announcement/emb).	
CC	or send an email to license@isb-sib.ch .	
CC	-----	
DR	EMBL; U75329; AAC51784.1; -	
DR	HSSP; F00763; IDPO.	
DR	MIM; 602060; -	
DR	PROSITE; PS00134; TRYPsin_HIS; 1.	
DR	PROSITE; PS00135; TRYPsin_SER; 1.	
DR	PROSITE; PS01209; LDLRA_1; 1.	
DR	PROSITE; PS00068; LDLRA_2; 1.	
DR	PFAM; PF00057; LDL_receptL_a; 1.	
DR	PFAM; PF00089; trypsin; 1.	
KW	Hydrolase; Serine protease; Transmembrane; Signal-anchor.	
FT	DOMAIN 1 84	
FT	TRANSMEM 85 105	
FT	POTENTIAL.	

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 2479)
Peoloni-Giacobino, A., Chen, H., Peltsch, M. C., Roessler, C. and Antonarakis, S. E.
TITLE Cloning of the TM6RS2 gene, which encodes a novel serine protease with transmembrane, LDLR, and SRCR domains and maps to 21q22.3
JOURNAL Genomics 44 (3), 309-320 (1997)
MEDLINE 97468144
REFERENCE 2 (bases 1 to 2479)
AUTHORS Peoloni-Giacobino, A., Chen, H. and Antonarakis, S. E.
TITLE Direct Submission
JOURNAL Submitted (17-OCT-1996) Medical Genetics, University of Geneva Medical School, 1 Michel-Servet Street, Geneva 1211, Switzerland
FEATURES
source
1. 2479
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
57..1535
/codon_start=1
/product="serine protease"
/protein_id="AAC51784.1"
/db_xref="GI:2507613"
/translation="MALNSGSPPAIPGPIYENHGYDPENPYPAOPTVVPVYEVHPAOY
YSPVPOYARVLTOASNPVYQPKSPSGTVCSTKALCTTLTFLVGAALAA
GLAKFMGSKNSGIECDSPGCIINPNSGDCGPNEDNRCVRLGPNFILLM
YSSORSMHPVCCDDNNENYRAACDMKMGKNGYSOGIVDSGSTFPMKLTSGN
VDYKRLHSDKSSKAVYSLRCLAGVNNSSROSRITYGESALGAMWQVSLHVO
NVAVCGSITTPETPIYTAACHVCEKPLNPHHTAPRAGILROSMFAGAGQONKVI
PNYDSFTKNDLILMKLQKRLFTNDLKPCLPCLPFLPCLPCLPCLPCLPCLPCLP
TSEVLAARVLIETQKNSRYVDNLITPAMICAGFLQGNVDSGSGPLVTNN
NIMWLGITSMGSCAKAYRPYGVYGVNMTTDMYIROMKANG"

BASE COUNT 578 a 650 c 677 g 574 t
ORIGIN

Query Match 95.28; Score 1655.2; DB 11; Length 2479;
Best Local Similarity 99.58; Pred. No. 0;
Matches 1660; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 56 gtcataatgaacattccagatacctatcattacccgattgctggtgataaagcgaagtgg 115
DB 1 CTCATATTGAACATTCCAGATACCTATCATCTGATGCTGTGATPAACGCAAGATGG 60
QY 116 cttagaactcagggctcaccacacagcattggaacttaactatgaataacatgataccaac 175
DB 61 CTTTGAACCTAGGGTCAACACCACTATGGAACCTTACTATGAAAACCATGGATACCAAC 120
QY 176 cggaaaaccctatcccgacagaccactgtgtgtcccaactgtcttaagagggtgacccgg 235
DB 121 CGGAAAACCCCTATCCGACACAGCCACTGTGCCACACTGTACAGAGGTGATCGCG 180
QY 236 ctcaagtaactcaggtcccccgtgtcccaagtaagcccgagggtctcttaagcagcttcca 295
DB 181 CTGAGTACTACCCCTCCCTGCGCCCAAGTACGCCCGAGGAGTCTGACGAGGCTTCCA 240
QY 296 accccgctgtgtgacgacagcccaatcccaatccggagaggtgtgacccccaagaagacta 355
DB 241 ACCCCGCTGTGTGACGACGAGCCAAATCCCATCCGGACAGTGTGACCTCAAAAGACTA 300
QY 356 agaaaagactgtgatacacttgacaccttgaggagaccttccgtgtggaggtgctgtggcgg 415
DB 301 AGAAAGACGCTGTGATACCTTGACCCGGGAGACTTCTCTGTGGAGAGCTCGCTGGCGG 360
QY 416 ctggcctaactctgaaagtcatgtggcagcaagtgtctccaactctggagtagagtgagact 475
DB 361 CTGGCCTACTCTGGAAGTTCATGGGCAAGTCTCTCCAACTCTGCGGTAAGTGGCGACT 420
QY 476 cctcaggttaactgatacaaccctctaactgtgtgatagtgctgtcaactgtcccccggc 535

DB 421 CCTACAGTACTCATCAACACCCCTTAACGTGGTGTATGAGCGGTGTACACTGCCCCGGCG 480
QY 536 gggaagagcaggaatctgt 595
DB 481 GGGAGGACGGAATCGGT 540
QY 596 catctcagaagaatctgt 655
DB 541 CATCTGAGAGAGTCTGT 600
QY 656 gggcggtctgcaaggacatggt 715
DB 601 GGGGGGCTCAGAGACATGGGCTATAGAAATATTTTACTGTACCAAGGAATAGTGG 660
QY 716 atgacagcagatcccaacagcttatagaactgaaccaggtgtgtgtgtgtgtgtgtgt 775
DB 661 ATGACACGGATCCACCAAGTTATGAACTGAACAACTGCCGCAATGTGATATCT 720
QY 776 ataaaaactgtacacacagtgatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 835
DB 721 ATAAAAACTGTACACAGT 780
QY 836 taagctgctggtgtcaactgtgaactcaagccgcagagcagagatgtgtgtgtgtgtgtgt 895
DB 781 TAGGCTGGGGGTCACCTTGAACCAAGCCGACAGACAGATGTGTGGCGGTGAGAGCG 840
QY 896 cgtctccgggt 955
DB 841 CGCTCCGGGGGCTGT 900
QY 956 gtagctcatcatcaaccccgagtgatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1015
DB 901 GAGGCTCATCATCACCCCGAGT 960
QY 1016 ttaacaatcaatgtgcatgtgacgagcattgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1075
DB 961 TTAAACATCCATGAGCATTTGAGCGGATTTGAGACATCTTATCATGTTC 1020
QY 1076 atggaagcagatataccaagt 1135
DB 1021 ATGAGCGCGGATGCCAAGTACAAAAGTATTTCTCTCAAAATTTATGATCCAAAGCCA 1080
QY 1136 agaaacaatgacatgt 1195
DB 1081 AGAACATGACATTTGCCCTGTATGAAGTGCAGAAAGCTGTGACTTTCAAGACTGTGTA 1140
QY 1196 aaccagtggt 1255
DB 1141 AACCATGT 1200
QY 1256 ccgggt 1315
DB 1201 CCGGTTGGGGGCGCCAGAGAGAAAGGAAGACCTTGAAGAGCTGAAAGCTGCCAAGG 1260
QY 1316 tgcctctcatgt 1375
DB 1261 TGCTTCTCATTTGAACACAGAGATGATGACAGAGATATGTCTATGACAACTTATCACAC 1320
QY 1376 cagcagatcatgt 1435
DB 1321 CAGCATGATGT 1380
QY 1436 gaaggcctgt 1495
DB 1381 GAGGCGCTGT 1440
QY 1496 gtctgt 1555
DB 1441 GTTCTGT 1500
QY 1556 actgattatcgacaaatgaaggcagcagctaatccaatgtgtgtgtgtgtgtgtgtgtgtgtgt 1615

FT	DOMAIN	240	350	CUB.
FT	DOMAIN	358	520	MAM.
FT	DOMAIN	540	650	CUB.
FT	DOMAIN	657	695	LID-RECEPTOR CLASS A 2.
FT	DOMAIN	694	787	SRR.
FT	ACT_SITE	841	841	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	892	892	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	987	987	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	LIPID	2	2	MYOSTATE (POTENTIAL).
FT	DISELFD	199	212	BY SIMILARITY.
FT	DISELFD	206	225	BY SIMILARITY.
FT	DISELFD	219	236	BY SIMILARITY.
FT	DISELFD	659	671	BY SIMILARITY.
FT	DISELFD	666	684	BY SIMILARITY.
FT	DISELFD	678	693	BY SIMILARITY.
FT	DISELFD	788	912	INTERCHAIN (BY SIMILARITY).
FT	DISELFD	826	842	BY SIMILARITY.
FT	DISELFD	926	993	BY SIMILARITY.
FT	DISELFD	957	972	BY SIMILARITY.
FT	DISELFD	983	1011	BY SIMILARITY.
FT	CARBONYD	116	116	POTENTIAL.
FT	CARBONYD	147	147	POTENTIAL.
FT	CARBONYD	170	170	POTENTIAL.
FT	CARBONYD	194	194	POTENTIAL.
FT	CARBONYD	233	233	POTENTIAL.
FT	CARBONYD	263	263	POTENTIAL.
FT	CARBONYD	264	264	POTENTIAL.
FT	CARBONYD	404	404	POTENTIAL.
FT	CARBONYD	456	456	POTENTIAL.
FT	CARBONYD	486	486	POTENTIAL.
FT	CARBONYD	519	519	POTENTIAL.
FT	CARBONYD	530	530	POTENTIAL.
FT	CARBONYD	646	646	POTENTIAL.
FT	CARBONYD	698	698	POTENTIAL.
FT	CARBONYD	722	722	POTENTIAL.
FT	CARBONYD	741	741	POTENTIAL.
FT	CARBONYD	762	762	POTENTIAL.
FT	CARBONYD	864	864	POTENTIAL.
FT	CARBONYD	903	903	POTENTIAL.
FT	CARBONYD	965	965	POTENTIAL.
FT	VARSPLIC	166	192	POTENTIAL.
FT	CONFLICT	808	808	MISSING (IN SHORT ISOFORM).
SO	SEQUENCE	1035 AA;	114887 MW;	R -> Y (IN REF. 3).
Query Match				
Best Local Similarity 24.4%; Score 95; DB 1; Length 1035;				
Matches 138; Conservative 36.3%; Pred.No.1.37e-211;				
Matches 138; Conservative 84; Mismatches 141; Indels 17; Gaps 16;				
Db	659 CKEDNFOCKD--GECIPLVNLCDGFPPHCKDSDEAHCYRLRNGTIDSSGLVOFRQSI-WH	716		
Oy	113 CSNSGIECDSSGTCITINSNMCDGVSHPGGEDENRCVRLY-GP-NFL-LQVYSQRKSMW	169		
Db	717 VACAENNMTQTSDDVVCOLLGLGTG-NSNSVPFVS-TGGGPVYNLTAP-NGSLILT-PSQQ	772		
Oy	170 PVGODDMNNTGRACRGMGTKNNFYISSQIGIVDDSGSTSFMKLTSGANDVIKKIYSDA	229		
Db	773 CLEDSLILLOCNKYKSCGKTLTOEVSPIRYGSDSREGAMPVVVALFYDDOYGASLVS	832		
Oy	230 CSSRAVYSLRCI--ACGVNLNSSMQS-RIVGEGALPGAMPQVLSLVHVNHWCGSIGIT	286		
Db	833 FDLVLSAHAHCYGRNME-PSRWKKAVLGLMAHSNLTSPQIEIRLLDQVLYNPHYKRKKNN	891		
Oy	287 FEWTYTAHAHV-EKPLPNPMHMTFAAGILRSQFMFGA-GYO-YEKYISHENDSKTKNN	343		
Db	892 DIAMMHLEMKNYNDYDIPICLPBENOVPFPRGCSINMGALYYOSTADVLDQADVPL	951		
Oy	344 DIALKLKPKLPFLFNDAVPCVLPNPGMALDPQLCLWISGKATETBKGTSEVLAQAAYLL	403		
Db	952 LSNKCCQOO-MPEYVITENNYCAGYEAGVGYSCOGSDSGPILMCOENNRMLLAGVTSEGYO	1010		
Oy	404 IETQCNSRIYYNDNLITPAMICAGELGVNDVSCGDGSGPLVTSKNNIMTWLIGDTSMGSG	463		
Db	1011 CALPNRGVYAVRPRTEMI	1030		

```

0y      464 CAAKAYRPGYGNMVFETDML 483
      11: 11111: 1 11:11
RESULT      3
ID ENTK_HUMAN STANDARD; PRT; 1019 AA.
AC P98073;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE ENTEROPEPTIDASE PRECURSOR (EC 3.4.21.9) (ENTEROKINASE).
PR PSS7 OR ENTK.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=DUDENUM;
RX MEDLINE: 95234679.
RA KITAMOTO Y., VELLE R.A., DONIS-KELLER H., SADLER J.E.;
RT "cDNA sequence and chromosomal localization of human enterokinase,
RL the proteolytic activator of trypsinogen.";
RN Biochemistry 34:4562-4568(1995).
RP [2]
RP SEQUENCE OF 749-1019 FROM N.A.
RC TISSUE=DUDENUM;
RX MEDLINE: 94329361.
RA KITAMOTO Y., YUAN X., WU Q., MCCOURT D.W., SADLER J.E.;
RT "Enterokinase, the initiator of intestinal digestion, is a mosaic
RL protease composed of a distinctive assortment of domains.";
RN Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
RP
CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
CC PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
CC A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
CC PROCARBOXYPEPTIDASES, AND PROELASTASES.
CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-|-ILE-7 BOND IN
CC TRYP-SINOGEN.
CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
CC MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC -1- TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.
CC -1- PM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CC CLEAVED BY A TRYPSIN-LIKE PROTEASE.
CC -1- DISEASE: DEFECTS IN PSS7 CAUSE LIFE-THREATENING INTESTINAL
CC MALABSORPTION CHARACTERIZED BY DIARRHEA AND FAILURE TO THRIVE.
CC -1- SIMILARITY: CONTAINS 2 LID-RECEPTOR CLASS A DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announcement/).
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U09860; AAC50138.1; -.
DR HSSP: P00763; IDPO.
DR MIM: 226200; -.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR PROSITE: PS0180; CUB; 2.
DR PROSITE: PS00740; MAM_1; 1.
DR PROSITE: PS00740; MAM_2; 1.
DR PROSITE: PS01209; LIDRA_1; 2.
DR PROSITE: PS00068; LIDRA_2; 2.
DR PFAM: PF00057; LID_recept_a; 2.
DR PFAM: PF00089; trypsin; 1.

```


misc_feature	15697..15808	/note="GRAL1 2 excellent exon, frame 0"
repeat_region	complement(15858..16136)	/rpt_family="Alu"
misc_feature	16365..16698	/note="GRAL1 2 excellent exon, frame 1"
repeat_region	17631..17863	/rpt_family="Alu"
repeat_region	complement(16891..19253)	/rpt_family="Alu"
repeat_region	complement(19424..19715)	/rpt_family="Alu"
misc_feature	20241..20286	/note="GRAL1 2 excellent exon, frame 0"
repeat_region	21081..21263	/rpt_family="MER5"
repeat_region	21861..22013	/rpt_family="MER20"
repeat_region	complement(22391..23728)	/rpt_family="THE1"
misc_feature	23779..23994	/note="GRAL 2 excellent exon, frame 0"
repeat_region	complement(24012..24087)	/rpt_family="MLT1"
misc_feature	24957..25094	/note="GRAL 2 excellent exon, frame 1"
repeat_region	25319..25585	/rpt_family="Alu"
misc_feature	27635..27825	/note="GRAL 2 excellent exon, frame 0"
repeat_region	28141..28430	/rpt_family="Alu"
misc_feature	28834..28966	/note="GRAL 2 excellent exon, frame 0"
misc_feature	29666..29929	/note="GRAL 2 excellent exon, frame 2"
repeat_region	29977..30000	/note="(GT)12"
misc_feature	31701..31844	/rpt_type="tandem
repeat_region	31701..31844	/note="GRAL1 2 excellent exon, frame 2"
repeat_region	complement(31932..32057)	/rpt_family="MER5"
repeat_region	32297..32320	/note="(T)24"
repeat_region	32997..33230	/rpt_type="tandem
repeat_region	33424..33584	/note="GRAL1 2 excellent exon, frame 0"
repeat_region	33976..34117	/note="GRAL 2 excellent exon, frame 0"
repeat_region	complement(34462..34558)	/rpt_family="19ger1"
repeat_region	complement(34577..34696)	/rpt_family="19ger1"
repeat_region	complement(35034..35111)	/rpt_family="MLT1"
repeat_region	complement(35713..36032)	/rpt_family="Alu"
repeat_region	36501..36803	/rpt_family="Alu"
repeat_region	36909..36950	/note="(GT)21"
repeat_region	37127..37266	/rpt_type="tandem
repeat_region	39543..39622	/note="GRAL 2 excellent exon, frame 2"
repeat_region	40597..40840	/rpt_family="Alu"

```

repeat_region    /note="GRAIL 2 excellent exon, frame 1"
                  /rpt_family="MER1"
                  complement(40977..41337)
repeat_region    /rpt_family="MER4"
                  complement(41617..41717)
repeat_region    /rpt_family="MER4"
                  complement(42881..43215)
repeat_region    /rpt_family="Alu"
                  43342..44219
repeat_region    /rpt_family="Alu"
                  complement(44724..45031)
repeat_region    /rpt_family="Alu"
                  complement(46007..46178)
repeat_region    /rpt_family="Alu"
                  46500..46725
repeat_region    /note="GRAIL 2 excellent exon, frame 2"
                  47500..47781
repeat_region    /rpt_family="Alu"
                  complement(49242..49309)
repeat_region    /note="GRAIL 2 excellent exon, frame 2"
                  49935..50077
repeat_region    /rpt_family="TBE1"
                  50080..51967
repeat_region    /rpt_family="WSTAR"
                  complement(50244..50543)
repeat_region    /rpt_family="Alu"
                  complement(51595..51645)
repeat_region    /note="GRAIL 2 excellent exon, frame 1"
                  51368..52329
repeat_region    /rpt_family="TBE1"
                  53080..55378
repeat_region    /rpt_family="Alu"
                  complement(56369..56511)
repeat_region    /note="GRAIL 2 excellent exon, frame 0"
                  complement(58621..58678)
repeat_region    /note="GRAIL 2 excellent exon, frame 0"
                  59339..59470
repeat_region    /rpt_family="MER20"
                  complement(59799..59910)
repeat_region    /note="GRAIL 2 excellent exon, frame 1"
                  14066 c 14518 g 15564 t

```

Query Match	10.2%;	Score 176.6;	DB 11;	Length 60904;
Best Local Similarity	93.0%;	Pred. No. 1.6e-36;		
Matches 185;	Conservative	0;	Mismatches 14;	Indels 0;
			Gaps	0;

Qy	988	acgcgcgcgcacacgcgctgtagaanaaacctctaaacaatccatgacatggatctggaaagcattcgcg	1047
Db	59991	actgagtcctctctctctcccccagacctttaaacaattccatggcatttgagacgcattttggc	59933
Qy	1048	ggagattcttgagacaactctctcatctctcatctgagccgcgatataccaagtagaanaaaagcatt	1107
Db	59931	gggatttttgagacaatacctttcatgtttcrtatgagaccggatnccaagtagaanaaaagcatt	59872
Qy	1108	tctcatccaataatgatgactcccaagaaccaagaacaaatgacatctgcgtatgaaagctgcg	1167
Db	59871	tcctcatccaataatgatgactcccaagaaccaagaacaaatgacatctgcgtatgaaagctgcg	59812
Qy	1168	aagcctctgactctcaacg	1186
Db	59811	aagcctctgactctcaacg	59793

RESULT	4		
HSMX1_5/C			
WPCOMMENT			
Sequence split into 7 fragments			
Fragment Name	Begin End Locus HSMX1.Accession AJ011928		
HSMX1_0	110000		
HSMX1_1	100001	210000	
HSMX1_2	200001	310000	
HSMX1_3	300001	410000	

KM Signal-anchor; Glycoprotein; Myristate; Hydrolase;
 KM Serine protease; Zymogen; Transmembrane; Repeat.
 FT CHAIN 52 117 NON-CATALYTIC H CHAIN (MINI CHAIN).
 FT CHAIN 118 799 NON-CATALYTIC H CHAIN (HEAVY CHAIN).
 FT CHAIN 800 1034 CATALYTIC L CHAIN (LIGHT CHAIN).
 FT TRANSMEM 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
 FT DOMAIN 197 238 LDL-RECEPTOR CLASS A 1.
 FT DOMAIN 240 349 CUB.
 FT DOMAIN 357 519 MAM.
 FT DOMAIN 357 519 CUB.
 FT DOMAIN 357 519 LDL-RECEPTOR CLASS A 2.
 FT DOMAIN 357 519 SRCR.
 FT ACT_SITE 693 786 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 840 840 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 891 891 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 986 986 MYRISTATE (POTENTIAL).
 FT LIPID 2 2
 FT DISULFID 199 212
 FT DISULFID 206 225
 FT DISULFID 219 236
 FT DISULFID 658 670
 FT DISULFID 665 683
 FT DISULFID 677 692
 FT DISULFID 787 911
 FT DISULFID 825 841
 FT DISULFID 925 992
 FT DISULFID 956 971
 FT DISULFID 982 1010
 FT DISULFID 116 116
 FT CARBOHYD 147 147
 FT CARBOHYD 170 170
 FT CARBOHYD 194 194
 FT CARBOHYD 283 283
 FT CARBOHYD 343 343
 FT CARBOHYD 350 350
 FT CARBOHYD 403 403
 FT CARBOHYD 455 455
 FT CARBOHYD 485 485
 FT CARBOHYD 518 518
 FT CARBOHYD 549 549
 FT CARBOHYD 645 645
 FT CARBOHYD 697 697
 FT CARBOHYD 701 701
 FT CARBOHYD 721 721
 FT CARBOHYD 740 740
 FT CARBOHYD 761 761
 FT CARBOHYD 804 804
 FT CARBOHYD 863 863
 FT CARBOHYD 902 902
 FT CARBOHYD 964 964
 SO SEQUENCE 1034 AA; 114776 MW; 243B6471 CRC32;

Query Match 23.5%; Score 862; DB 1; Length 1034;
 Best Local Similarity 35.0%; Pred. No. 3,79e-202;
 Matches 133; Conservative 89; Mismatches 141; Indels 17; Gaps 16;

OY 344 DIALMLKLOKPLTFNDLVKPYCLPNPBMLOPEOLCWMISNGATEKCKTSEVLAARVLL 403
 DB 951 LSNEKCGOO-MPEVNTENMMACAGEEGIDSCGDSGGPLMCLENNRWLLAGVTSFGYQ 1009
 OY 404 IETORCSRRYIDLLTPAMICAGFLOGNDVSCGDSGGPLVTSKRNIMWLLIGDTSWNGS 463
 DB 1010 CALPNRPQYVAPKTEWMI 1029
 OY 464 CARKRPQYVGNVVFETDWI 483

RESULT 5
 ID ENTK_MOUSE STANDARD; PRT; 1069 AA.
 AC P97435;
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE ENTEROPEPTIDASE (EC 3.4.21.9) (ENTEROKINASE).
 GN PRSS7 OR ENTK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Rodentia; Sclurognath; Muridae; Murinae; Mus.
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN-C57BL/6; TISSUE-DUODENUM;
 RA YUAN X., LU D., RUBIN D.C., PUNG C.Y.M., SADLER J.E.;
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
 CC PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
 CC A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
 CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
 CC PROCARBOXYPEPTIDASES, AND PROELASTASES (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
 CC TRYPSINOGEN.
 CC -1- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
 CC MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
 CC -1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
 CC CLEAVED BY A TRYPSIN-LIKE PROTEASE (BY SIMILARITY).
 CC -1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL: U73378; AAB37317.1; -
 DR HSPB; P00763; IDPO.
 DR MGD; MGI:1197523; PRSS7.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR PROSITE; PS01180; CUB; 2.
 DR PROSITE; PS00740; MAM_1; 1.
 DR PROSITE; PS00060; MAM_2; 1.
 DR PROSITE; PS01209; LDLR_1; 2.
 DR PROSITE; PS50068; LDLR_2; 2.
 DR PFAM; PF00057; Idl_recept_a; 2.
 DR PFAM; PF00089; trypsin; 1.
 DR PFAM; PF00431; CUB; 2.
 DR PFAM; PF00530; SRCR; 1.
 DR PFAM; PF00629; MAM; 1.
 DR PFAM; PF01390; SEA; 1.
 DR Signal-anchor; Glycoprotein; Myristate; Hydrolase;
 KM Serine protease; Zymogen; Transmembrane; Repeat.

KEYWORDS `prekallikrein`.


```

FT CARBOHYD 127 127
FT CARBOHYD 308 308
FT CARBOHYD 396 396
FT CARBOHYD 453 453
FT CARBOHYD 494 494
FT ACT_SITE 434 434
FT ACT_SITE 483 483
FT ACT_SITE 578 578
FT DISULFID 21 104
FT DISULFID 47 77
FT DISULFID 51 57
FT DISULFID 111 194
FT DISULFID 137 166
FT DISULFID 141 147
FT DISULFID 201 284
FT DISULFID 227 256
FT DISULFID 231 237
FT DISULFID 292 375
FT DISULFID 318 347
FT DISULFID 322 328
FT DISULFID 340 345
FT DISULFID 383 503
FT DISULFID 419 435
FT DISULFID 517 584
FT DISULFID 548 563
FT DISULFID 574 602
SQ SEQUENCE 638 AA; 71369 MW; DAF630DA CRC32;

Query Match 20.7%; Score 762; DB 1; Length 638;
Best Local Similarity 42.9%; Pred. No. 1,156-173;
Matches 109; Conservative 51; Mismatches 84; Indels 10; Gaps 9;

Db 371 SLRCLNCDNSVCTKSTRIVGCTNSKEMWPMOVSLTNRHLCGSLGHOMVL 430
Oy 237 SLR-CICGVNLSSROS-RIVGESALPGAMPVOSLHV-NV--HVCSSITTPENTV 291
Db 431 TAACFCGLPLQDVR-1-YGIILNLSIDITKDPFSQKEIILHONYKSEGNHDLALIK 488
Oy 292 TAAACVEK-PLNNPMHMTAFAGILROSMFAGY-OVEKYSHPNDOSKTRNDIALMK 349
Db 489 LQAPLNTTEFOKPICLPSKDTSTIYNCWVTGSGFSEKEKEIIONILOKVNPITYNSEC 548
Oy 350 LQKPLTFENDLVKPVCLPFGMLOPEOLCMTSGWATEKEKTSVLNAKVLLETORC 409
Db 549 CKRRO-DKTIORVAVCGYKEGKDGACKDGGSGPLVCKHNMNRLVGTSGEGCARRO 607
Oy 410 NSRYVDLITPAMICAGFLOGVNDSCGDSGGLVTSKNNIMWLGDTSGSCAKAYR 469
Db 608 PGVYTKVAEYMDWI 621
Oy 470 PGYGVNVAFTDWT 483

RESULT 7
ID KAL_RAT STANDARD; PRT: 638 AA.
AC P14272;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE PLASMA KALLIKREIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKREIN)
DE (KININOGENIN) (FLETCHER FACTOR).
OS PK.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE: 91129236.
RA BEAUBIEN G., ROSINSKI-CHUPIN I., MATTEI M.-G., MBIKAY M., CHRETIEN M.,
RA SEIDAH N.G.;
RL "Gene structure and chromosomal localization of plasma kallikrein.";
RL Biochemistry 30:1628-1635(1991).
[2]

```

```

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RA MEDLINE: 90091743.
RA SEIDAH N.G., LADENHEIM R., MBIKAY M., HAMELIN J., LUTFALLA G.,
RA ROUGEON F., LAZURE C., CHRETIEN M.;
RT "The cDNA structure of rat plasma kallikrein.";
RL DNA 8:563-574(1989).
CC -1- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT
CC ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING
CC TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM
CC HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
CC SYSTEM BY CONVERTING PRORENIN INTO RENIN.
CC -1- SUBUNIT: THE ZYMOGEN IS ACTIVATED BY FACTOR XIIa, WHICH CLEAVES
CC THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,
CC AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE
CC CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
CC -1- SIMILARITY: CONTAINS 4 APPLE DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY. BELONGS TO THE PLASMA KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.sib.ch/announce/
CC or send an email to license@isb.sib.ch).
CC -----
DR EMBL: M63357; AAA74563.1; -.
DR EMBL: M63358; AAA74563.1; JOINED.
DR EMBL: M63346; AAA74563.1; JOINED.
DR EMBL: M63347; AAA74563.1; JOINED.
DR EMBL: M63349; AAA74563.1; JOINED.
DR EMBL: M63350; AAA74563.1; JOINED.
DR EMBL: M63351; AAA74563.1; JOINED.
DR EMBL: M63352; AAA74563.1; JOINED.
DR EMBL: M63353; AAA74563.1; JOINED.
DR EMBL: M63354; AAA74563.1; JOINED.
DR EMBL: M63355; AAA74563.1; JOINED.
DR EMBL: M63356; AAA74563.1; JOINED.
DR EMBL: M30282; AAA41463.1; -.
DR EMBL: M58590; AAA42069.1; -.
DR PIR: A39180; KORTPL.
DR HSP: P00750; IRTF.
DR PROSITE: PS00134; TRYPsin_HIS. 1.
DR PROSITE: PS00135; TRYPsin_SER. 1.
DR PROSITE: PS00495; APPLE. 4.
DR PRAM: PF00024; apple. 4.
DR PRAM: PF00089; trypsin. 1.
DR Hydrolase; Serine protease; Glycoprotein; Plasma; Zymogen; Signal;
KW Fibrinolysis; Blood coagulation; Inflammatory response; Layer;
KM Duplication; Bradykinin.
FT SIGNAL 1 19
FT CHAIN 20 390
FT CHAIN 391 638
FT REPEAT 20 105
FT REPEAT 110 195
FT REPEAT 200 285
FT REPEAT 291 376
FT DOMAIN 389 621
FT CARBOHYD 127 127
FT CARBOHYD 308 308
FT CARBOHYD 396 396
FT CARBOHYD 453 453
FT CARBOHYD 494 494
FT ACT_SITE 434 434
FT ACT_SITE 483 483
FT ACT_SITE 578 578
FT DISULFID 21 104
FT DISULFID 47 77
FT DISULFID 51 57
FT DISULFID 111 194
FT DISULFID 137 166
FT DISULFID 141 147

PLASMA KALLIKREIN HEAVY CHAIN.
PLASMA KALLIKREIN LIGHT CHAIN.
APPLE 1.
APPLE 2.
APPLE 3.
APPLE 4.
CATALYTIC.
PROBABLE.
PROBABLE.
PROBABLE.
PROBABLE.
PROBABLE.
CHARGE RELAY SYSTEM.
CHARGE RELAY SYSTEM.
CHARGE RELAY SYSTEM.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.

```

```

GENERAL INFORMATION:
APPLICANT: Morrissey, James H.
APPLICANT: Comp, Philip C.
TITLE OF INVENTION: Truncated Tissue Factor and FvIII or
TITLE OF INVENTION: FvIII Activator for Blood Coagulation
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESS: Richards, Medlock & Andrews
STREET: 1201 Elm Street, Suite 4500
CITY: Dallas
STATE: Texas
COUNTRY: US
ZIP: 75270-2197
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/04493
FILING DATE: 19930512
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/882202
FILING DATE: 13-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021615
FILING DATE: 19-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Trujillo, Doreen Y.
REGISTRATION NUMBER: 35,719
REFERENCE/DOCKET NUMBER: OMRF B34290CIPC/PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 214-939-4500
TELEFAX: 214-939-4600
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1440 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Blood
FEATURE:
NAME/KEY: CDS
LOCATION: 36..1433
OTHER INFORMATION: /product= "Tissue Factor"
OTHER INFORMATION: /note= "Coding portion of human factor VIII cDNA"
OTHER INFORMATION: /citation= ([1])
PCT-US93-04493-3
Query Match 5.5%; Score 95.4; DB 6; Length 1440;
Best Local Similarity 47.9%; Pred. No. 9.4e-17;
Matches 353; Conservative 0; Mismatches 366; Indels 18; Gaps 2.
QY 865 cgcacagacagagattgtgtggcgagagagcgctcccggtggtcctgagccctggcagatgc 924
Db 660 CCCCAGGCGCCGAGATGTGTGGGGGCAAGGTGTCCCCAAAGGGAGTGTCCATGCGCGGTC 719
QY 925 agccgcacgtccagaaagctccagctgtgtggagaggtccatcatcaaccgccagatgac 984
Db 720 CTGTGTGTGGTAAGAGACTCATGTTGTGTGGGGGACCCCTGATCAACACCTCTGTGGGTG 779
QY 985 gtacacgcgcgcacatgcgtggaaaaaacctcttaacaatccatccatctgcatgtgacgcat 1044
Db 780 GTCTCGCGCGCCCACTGTTTCGACAAATCAAGAACTGAGAGAACTGATCGCGGTGTG 839
QY 1045 gctggagatttgagacaatcttcatgttctctatgtgagccgagataccagtagaaaaatg 1104

```

```

Db      840 GGGAGACGACACTCTAGCAGCAGCACAGCGGGGATGTAGCAGAGCCGGCGGGTGGCCAGSTC   899
        |         |         |         |         |         |         |         |
Qy      1105 attctcattcaaatatayactccaaagacaagaacaaatgacattggctaytaagtgc   1164
        ||         ||         ||         ||         ||         ||         ||
Db      900 ATCATCCCCAGCAGCACTGCCGCCGGGCACACCAACGACGACATCGCCTTCCTCCGCTG   939
        |         |         |         |         |         |         |         |
Qy      1165 cagaagacttgcttcttcaaagcaactgaagaaacaaagtgctgtccccaa-----c   1215
        |         |         |         |         |         |         |         |
Db      960 CACACAGCCCTGTGCTCTCACTGACCACTGTGTGCCCTCTTGCTGTGCCGAAGCAGCTTC   1019
        |         |         |         |         |         |         |         |
Qy      1216 ccaggacatgatgtctgacgccaaagacagctctgtgattccgggtgggaggaccaggag   1275
        |         |         |         |         |         |         |         |
Db      1020 TCTGAGAGAGACGCTGGCCTTTCTGTGCCCTTCTCATTTGTGTACGGCGCTGGGGCACCTGCTG   1079
        |         |         |         |         |         |         |         |
Qy      1276 gagaagaaggaaagaccttcagaagtgtctgaaacgtlccaaagtgtctctcatgtagaacag   1335
        |         |         |         |         |         |         |         |
Db      1080 GACCCTGGCCGACAGCGGCCCGTAGAGCTCATGTGAGCTCAACGTCGCCCGGCTATGACCCAG   1139
        |         |         |         |         |         |         |         |
Qy      1336 agatgc-----aacagcgatatgtctatgnaacactgatacacacagccatgac     1386
        |         |         |         |         |         |         |         |
Db      1140 GATCGCCTGAGCAGATGACACGAAAGGTGGAGACTCCCAATATATACGAGTACTGTTC   1199
        |         |         |         |         |         |         |         |
Qy      1387 tgtagcgactcctctcgaggaggaaagcgtcgaattcttcgcaagtgtagcagtggaggagcctctg   1446
        |         |         |         |         |         |         |         |
Db      1200 TGTCGCCGCTACTTCGATGTGGCACCAGAAGACTCTCTGMAAGGGGAGACAGTGAAGCCCACAT   1259
        |         |         |         |         |         |         |         |
Qy      1447 gtccactcgagaagaacaatatctgtgtgctgatatgggatacaagctggggttcgtgcgt   1506
        |         |         |         |         |         |         |         |
Db      1260 GCACACCCATACCGGGGACAGCTGTAACGTACAGCGGCATCGTCAGCTGGGGGCCAGGGCTGC   1319
        |         |         |         |         |         |         |         |
Qy      1507 gccaaagcttaagaccagaagatgtctgggaatgtgatatctaagcagctggattat   1566
        |         |         |         |         |         |         |         |
Db      1320 GCAACCGCTGGGACCATTGTGGGGGTGTACACAGGGGTCTCCCACTGATGAGTGGCTGCA   1379
        |         |         |         |         |         |         |         |
Qy      1567 cgacaatgaagggcaga   1583
        |         |         |         |         |         |         |         |
Db      1380 AAGCTCATGCCCTCACA   1396
        |         |         |         |         |         |         |         |

RESULT 11
US-08-475-845-1
Sequence 1, Application US/08475845
Patent No. 5788965
GENERAL INFORMATION:
APPLICANT: Berkner, Kathleen L.
APPLICANT: Petersen, Lars C.
APPLICANT: Hart, Charles E.
APPLICANT: Hedner, Ulla
APPLICANT: Bregengaard, Claus
TITLE OF INVENTION: Modified Factor VII
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Kourile and Crew
STREET: One Market Plaza, Steuart Street Tower
CITY: San Francisco
STATE: CA
COUNTRY: U.S.A.
ZIP: 94105-1492
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,845
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/327,690
FILING DATE: 24-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

```

RESULT	9	STANDARD:	PRT:	625 AA.
ID	AC	FAIL HUMAN		
DT	23-OCT-1986	(Rel. 02, Created)		
DT	23-OCT-1986	(Rel. 02, Last sequence update)		
DT	15-DEC-1999	(Rel. 39, Last annotation update)		
DE	COAGULATION FACTOR XI	PRECURSOR (EC 3.4.21.27) (PLASMA THROMBOPLASTIN ANTECEDENT) (PIA).		
DE	FN1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
OC	Eutheria; Primates; Catarrhini; Homnidae; Homo.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 86243360.			
RA	FUJIKAWA K., CHUNG D.W., HENDRICKSON J.E., DAVIE E.W.;			
RT	"Amino acid sequence of human factor XI, a blood coagulation factor with four tandem repeats that are highly homologous with plasma prekallikrein.";			
RL	Biochemistry 25:2417-2424(1986).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 88107663.			
RA	ASAKAI R., DAVIE E.W., CHUNG D.W.;			
RT	"Organization of the gene for human factor XI.";			
RL	Biochemistry 26:7221-7228(1987).			
RN	[3]			
RP	PARTIAL SEQUENCE, AND DISULFIDE BONDS.			
RX	MEDLINE; 91152017.			
RA	MCWULLEN B.A., FUJIKAWA K., DAVIE E.W.;			
RT	"Location of the disulfide bonds in human coagulation factor XI: the presence of tandem apple domains.";			
RL	Biochemistry 30:2056-2060(1991).			
RN	[4]			
RP	VARIANT LEU-301.			
RX	MEDLINE; 90046656.			
RA	ASKAKI R., CHUNG D.W., RATNOFF O.D., DAVIE E.W.;			
RT	"Factor XI (plasma thromboplastin antecedent) deficiency in Ashkenazi Jews is a bleeding disorder that can result from three types of point mutations.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 86:7667-7671(1989).			
RN	[5]			
RP	VARIANT LEU-301.			
RX	MEDLINE; 92190478.			
RA	MEIJERS J.C., DAVIE E.W., CHUNG D.W.;			
RT	"Expression of human blood coagulation factor XI: characterization of the defect in factor XI type III deficiency.";			
RL	Blood 79:1435-1440(1992).			
CC	-1- FUNCTION: FACTOR XI TRIGGERS THE MIDDLE PHASE OF THE INTRINSIC PATHWAY OF BLOOD COAGULATION BY ACTIVATING FACTOR IX.			
CC	-1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF ARG-1-ALA AND ARG-1-VAL BONDS IN FACTOR IX TO FORM FACTOR IXA.			
CC	-1- SUBUNIT: MONOMER, LINKED BY A DISULFIDE BOND. AFTER ACTIVATION THE HEAVY AND LIGHT CHAINS ARE ALSO LINKED BY A DISULFIDE BOND.			
CC	-1- PMW: ACTIVATED BY FACTOR XIa (OR XIi), WHICH CLEAVES EACH POLYPEPTIDE AFTER ARG-387 INTO THE LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE, AND THE HEAVY CHAIN, WHICH ASSOCIATES WITH HIGH MOLECULAR WEIGHT (HMW) KININAGEN.			
CC	-1- DISEASE: DEFECTS IN FII ARE A CAUSE OF A BLOOD COAGULATION ANOMALY (ROSENTHAL SYNDROME) OCCURRING IN HIGH FREQUENCY IN ASHKENAZI JEWS.			
CC	-1- SIMILARITY: CONTAINS 4 APPLE DOMAINS.			
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE THROMBIN FAMILY. BELONGS TO THE PLASMA KALLIKREIN SUBFAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-slb.ch/announce/ or send an email to license@isb-slb.ch).			
CC	-----			

DR	EMBL; M13142;	AAAS2487.1;	-.
DR	EMBL; M20218;	AAAS1985.1;	-.
DR	EMBL; M18296;	AAAS1985.1;	JOINED.
DR	EMBL; M21184;	AAAS1985.1;	JOINED.
DR	EMBL; M18298;	AAAS1985.1;	JOINED.
DR	EMBL; M18299;	AAAS1985.1;	JOINED.
DR	EMBL; M18300;	AAAS1985.1;	JOINED.
DR	EMBL; M18301;	AAAS1985.1;	JOINED.
DR	EMBL; M18302;	AAAS1985.1;	JOINED.
DR	EMBL; M18303;	AAAS1985.1;	JOINED.
DR	EMBL; M18304;	AAAS1985.1;	JOINED.
DR	EMBL; M19417;	AAAS1985.1;	JOINED.
DR	EMBL; M20217;	AAAS1985.1;	JOINED.
DR	PIR; A27431;	KFHU1.	
DR	HSSP; P00763;	IDPO.	
DR	MIM; 134540;	-.	
DR	MIM; 264900;	-.	
DR	PROSITE; PS00134;	TRYPSIN_HIS. 1.	
DR	PROSITE; PS00135;	TRYPSIN_SER. 1.	
DR	PROSITE; PS00495;	APPLE. 4.	
DR	Pfam; PF00089;	trypsin. 1.	
KW	Hydrolase; Serine protease;	Glycoprotein; Plasma:	Blood coagulation;
RK	Duplication; Signal;	Disease mutation.	
FT	SIGNAL	1	18
FT	CHAIN	19	387
FT	CHAIN	388	625
FT	REPEAT	19	104
FT	REPEAT	109	194
FT	REPEAT	199	284
FT	REPEAT	290	375
FT	DOMAIN	384	625
FT	CARBOHYD	90	90
FT	CARBOHYD	126	126
FT	CARBOHYD	353	353
FT	CARBOHYD	450	450
FT	CARBOHYD	491	491
FT	ACT_SITE	431	431
FT	ACT_SITE	480	480
FT	ACT_SITE	575	575
FT	DISULFID	20	103
FT	DISULFID	29	29
FT	DISULFID	46	76
FT	DISULFID	50	56
FT	DISULFID	110	193
FT	DISULFID	136	165
FT	DISULFID	140	146
FT	DISULFID	200	283
FT	DISULFID	226	255
FT	DISULFID	230	236
FT	DISULFID	291	374
FT	DISULFID	317	346
FT	DISULFID	321	327
FT	DISULFID	339	339
FT	DISULFID	380	500
FT	DISULFID	416	432
FT	DISULFID	514	581
FT	DISULFID	545	560
FT	DISULFID	571	599
FT	VARIANT	301	301
FT	CONFLICT	226	226
FT	SEQUENCE	625 AA;	70109 MM;
DB	Query Match	18.9%;	Score 694;
DB	Best Local Similarity	42.6%;	DB 1;
Matches	100;	Conservative	43;
QY	23	SRVGGESALGCAHPMQVSLHVN-V--HYCGGSITTEPMVLTAACHCEKRLNPPMHMTA	Length 625; Pred. No. 1,96e-154; Indels 6; Gaps 5

APPLICANT: Hart, Charles E.
 APPLICANT: Petersen, Lars C.
 APPLICANT: Hedner, Ulla
 APPLICANT: Rasmussen, Mirella E.
 TITLE OF INVENTION: Modified Factor VII
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: Zymogenetics, Inc.
 STREET: 1201 Eastlake Avenue East
 CITY: Seattle
 STATE: WA
 COUNTRY: USA
 ZIP: 98102
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentln Release #1.24
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/871,003
 FILING DATE:
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 CLASSIFICATION: 514
 ATORNEY/AGENT INFORMATION:
 NAME: Sawislak, Deborah A
 REGISTRATION NUMBER: 37,438
 REFERENCE/DOCKET NUMBER: 90-07C7
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2422 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 HYPOTHETICAL: N
 ANTI-SENSE: N
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 28..1420
 OTHER INFORMATION: /codon_start= 28
 OTHER INFORMATION: /product= "Factor VII"
 US-08-871-003-1

DR PFAM: PF00051; kringle; 5.
 DR PFAM: PF00089; trypsin; 1.
 KM Hydrolase: Serine protease; Plasma: Glycoprotein; Fibrinolysis;
 KM Tissue remodeling; Blood coagulation; Kringle; Zymogen; Signal.
 FT SIGNAL 1
 FT CHAIN 20 812 PLASMINOGEN.
 FT CHAIN 20 581 PLASMINOGEN HEAVY CHAIN A.
 FT PEPTIDE 20 97 ACTIVATION PEPTIDE.
 FT CHAIN 98 581 PLASMINOGEN SHORT FORM OF CHAIN A.
 FT CHAIN 98 7436 ANGIOSTATIN.
 FT CHAIN 582 812 PLASMINOGEN LIGHT CHAIN B.
 FT DOMAIN 103 181 KRINGLE 1.
 FT DOMAIN 184 262 KRINGLE 2.
 FT DOMAIN 275 352 KRINGLE 3.
 FT DOMAIN 377 454 KRINGLE 4.
 FT DOMAIN 481 560 KRINGLE 5.
 FT DOMAIN 582 812 CATALYTIC.
 FT ACT_SITE 624 624 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 667 667 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 762 762 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 49 73 BY SIMILARITY.
 FT DISULFID 53 61 BY SIMILARITY.
 FT DISULFID 103 181 BY SIMILARITY.
 FT DISULFID 124 164 BY SIMILARITY.
 FT DISULFID 152 176 BY SIMILARITY.
 FT DISULFID 185 262 BY SIMILARITY.
 FT DISULFID 188 316 BY SIMILARITY.
 FT DISULFID 206 245 BY SIMILARITY.
 FT DISULFID 224 257 BY SIMILARITY.
 FT DISULFID 234 352 BY SIMILARITY.
 FT DISULFID 296 335 BY SIMILARITY.
 FT DISULFID 324 347 BY SIMILARITY.
 FT DISULFID 377 454 BY SIMILARITY.
 FT DISULFID 398 437 BY SIMILARITY.
 FT DISULFID 426 449 BY SIMILARITY.
 FT DISULFID 481 560 BY SIMILARITY.
 FT DISULFID 502 543 BY SIMILARITY.
 FT DISULFID 531 555 BY SIMILARITY.
 FT DISULFID 568 687 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 578 586 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 609 625 BY SIMILARITY.
 FT DISULFID 701 768 BY SIMILARITY.
 FT DISULFID 731 747 BY SIMILARITY.
 FT DISULFID 758 786 BY SIMILARITY.
 SQ SEQUENCE 812 AA: 90846 KM; 6C120F7A CRC32;
 Query Match 17.3%; Score 634; DB 1; Length 812;
 Best Local Similarity 41.4%; Pred. No. 1,41e-137;
 Matches 111; Conservative 51; Mismatches 93; Indels 13; Gaps 12;

AC P80010;
 DR 01-NOV-1991 (Rel. 20, Created)
 DR 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE PLASMINOGEN (EC 3.4.21.7) (FRAGMENT).
 GN PLG.
 OS Equus caballus (Horse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
 OC Eutheria; Perissodactyla; Equidae; Equus.
 RN [1]
 RP SEQUENCE.
 RC TISSUE-PLASMA;
 RX MEDLINE; 92052077.
 RA SCHALLER J., STRAUB C., KAEMPFER U., RICKLI E.E.;
 RT "Complete amino acid sequence of equine miniplasminogen";
 RL Protein Seq. Data Anal. 4:69-74(1991).
 CC -1- FUNCTION: PLASMIN DISSOLVES THE FIBRIN OF BLOOD CLOTS AND ACTS AS
 A PROTEOLYTIC FACTOR IN A VARIETY OF OTHER PROCESSES INCLUDING
 EMBRYONIC DEVELOPMENT, TISSUE REMODELING, TUMOR INVASION,
 AND INFLAMMATION; IN OVULATION IT WEAKENS THE WALLS OF THE
 GRAAFIAN FOLLICLE. IT ACTIVATES THE UKRINASE-TYPE PLASMINOGEN
 ACTIVATOR. COLLAGENASES AND SEVERAL COMPLEMENT ZYMOGENS, SUCH
 AS C1 AND C5, IT CLEAVES FIBRIN, FIBRONECTIN, THROMBOSPONDIN,
 LAMININ AND VON WILLEBRAND FACTOR.
 CC -1- ENZYME REGULATION: CONVERTED INTO PLASMIN BY PLASMINOGEN
 ACTIVATORS. BOTH PLASMINOGEN AND ITS ACTIVATOR BEING BOUND TO
 FIBRIN. ACTIVATED WITH CATALYTIC AMOUNTS OF STREPTOKINASE.
 CC -1- MISCELLANEOUS: PLASMIN IS INACTIVATED BY ALPHA-2-ANTIPLASMIN
 IMMEDIATELY AFTER DISSOCIATION FROM THE CLOT.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY. BELONGS TO THE PLASMINOGEN SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS AT LEAST 1 KRINGLE REGIONS.
 CC PIR: S17527; S17527.
 DR HSSP: P00747; SHRG.
 DR PROSITE: P500021; KRINGLE_1; 1.
 DR PROSITE: P500134; TRYPSIN_HIS; 1.
 DR PROSITE: P500135; TRYPSIN_SER; 1.
 DR PROSITE: P500070; KRINGLE_2; 1.
 DR PFAM: PF00051; kringle; 1.
 DR PFAM: PF00089; trypsin; 1.
 KM Hydrolase; Serine protease; Plasma: Glycoprotein; Fibrinolysis;
 KM Tissue remodeling; Blood coagulation; Kringle; Zymogen.
 FT NON_TER 1 1
 FT CHAIN <1 108 HEAVY CHAIN A.
 FT CHAIN 109 338 LIGHT CHAIN B.
 FT DOMAIN 9 88 KRINGLE 5.
 FT DISULFID 109 338 CATALYTIC.
 FT DISULFID 9 88 BY SIMILARITY.
 FT DISULFID 30 71 BY SIMILARITY.
 FT DISULFID 59 83 BY SIMILARITY.
 FT DISULFID 95 213 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 105 113 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 135 151 BY SIMILARITY.
 FT DISULFID 227 294 BY SIMILARITY.
 FT DISULFID 257 273 BY SIMILARITY.
 FT DISULFID 264 312 BY SIMILARITY.
 FT ACT_SITE 150 150 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 193 193 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 288 288 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT SITE 157 157 STREPTOKINASE-BINDING SITE (PROBABLE).
 FT SITE 191 191 STREPTOKINASE-BINDING SITE (PROBABLE).
 FT SITE 269 269 STREPTOKINASE-BINDING SITE (PROBABLE).
 FT SITE 282 282 SITE OF SUBSTRATE SPECIFICITY (BY SIMILARITY).
 SQ SEQUENCE 338 AA: 37132 KM; 81DD0C34 CRC32;
 Query Match 16.8%; Score 617; DB 1; Length 338;
 Best Local Similarity 38.2%; Pred. No. 8.01e-133;
 Matches 97; Conservative 66; Mismatches 79; Indels 12; Gaps 12;

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

Msearch.p protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Apr 19 23:28:21 2000; Maspar time 27.50 Seconds
842,259 Million cell updates/sec
Tabular output not generated.

Title: >US-09-323-597-2
Description: (1-491) from US09323597.pep
Sequence: 1 MALNGSPPAICPYENHGY.....VYGNVVFWDWIRQMRADG 491
Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: plr62
1:plr1 2:plr2 3:plr3 4:plr4

Statistics: Mean 47.373; Variance 79.961; scale 0.592

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description	Pred. No.
1	895	24.4	1035 1 A43090	enteropeptidase (EC 3	3.16e-180
2	890	24.2	1019 1 A56318	enteropeptidase (EC 3	5.25e-179
3	862	23.5	1034 1 A53663	enteropeptidase (EC 3	3.55e-172
4	762	20.7	638 1 KOHUP	plasma kallikrein (EC	6.82e-148
5	740	20.1	638 1 KORPL	plasma kallikrein (EC	1.42e-142
6	733	20.0	1113 2 JEO315	low-density lipoprote	6.93e-141
7	720	19.6	638 1 KOMSP	plasma kallikrein (EC	9.45e-138
8	694	18.9	625 1 KRHU1	coagulation factor XI	1.71e-131
9	680	18.5	417 1 S00845	hepsin (EC 3.4.21.-)	3.96e-128
10	652	17.8	416 1 S33777	hepsin (EC 3.4.21.-)	2.04e-121
11	634	17.3	812 1 PLMS	plasmin (EC 3.4.21.7)	4.10e-117
12	617	16.8	455 2 A61545	plasmin (EC 3.4.21.7)	4.65e-113
13	613	16.7	790 1 PIRG	plasmin (EC 3.4.21.7)	4.18e-112
14	608	16.6	810 2 B30848	plasmin (EC 3.4.21.7)	6.48e-111
15	597	16.3	810 1 PLHU	plasmin (EC 3.4.21.7)	1.12e-110
16	597	16.3	810 1 PLHU	plasmin (EC 3.4.21.7)	2.68e-108
17	583	15.9	267 2 S40006	trypsin (EC 3.4.21.4)	5.65e-105
18	582	15.8	273 2 A47246	trypsin (EC 3.4.21.5)	9.75e-105
19	577	15.7	276 2 A38654	mast cell proteinase	1.49e-103
20	572	15.6	275 2 C35863	trypsin (EC 3.4.21.5)	2.29e-102
21	572	15.6	275 2 C35863	trypsin (EC 3.4.21.5)	2.29e-102
22	572	15.6	275 2 A35863	trypsin (EC 3.4.21.5)	2.29e-102

24	568	15.5	421 1 S11674	acrosin (EC 3.4.21.10)	2.02e-101
25	570	15.5	810 2 I46260	plasmin (EC 3.4.21.7)	6.80e-102
26	569	15.5	812 1 PLBO	plasmin (EC 3.4.21.7)	1.17e-101
27	564	15.4	263 2 A21195	chymotrypsin (EC 3.4.	1.79e-100
28	564	15.4	343 1 A57014	prolactin (EC 3.4.21.	1.79e-100
29	561	15.3	274 2 JC4171	trypsin (EC 3.4.21.5)	9.16e-100
30	558	15.2	245 1 KYBOA	chymotrypsin (EC 3.4.	4.69e-99
31	557	15.2	275 2 S40007	trypsin (EC 3.4.21.4)	8.08e-99
32	555	15.1	245 1 KYBOB	chymotrypsin (EC 3.4.	2.40e-98
33	554	15.1	263 2 A31299	chymotrypsin (EC 3.4.	4.13e-98
34	546	14.9	264 2 I38136	chymotrypsin-like pro	3.19e-96
35	548	14.9	274 2 S35339	trypsin (EC 3.4.21.4)	1.08e-96
36	548	14.9	275 2 S40005	trypsin (EC 3.4.21.4)	1.85e-96
37	547	14.9	415 1 A34170	hepatocyte growth fac	3.19e-96
38	546	14.9	655 1 A46688	acrosin (EC 3.4.21.10)	9.45e-96
39	544	14.8	436 2 JX0172	trypsin (EC 3.4.21.4)	8.28e-95
40	540	14.7	277 2 S35340	chymotrypsin (EC 3.4.	2.45e-94
41	538	14.6	263 1 KYRTB	chymotrypsin (EC 3.4.	2.45e-94
42	538	14.6	263 2 S47537	protein C (activated)	1.87e-92
43	530	14.4	456 1 KXBO	serine proteinase st	3.22e-92
44	529	14.4	786 1 A47547	apolipoprotein(a) (EC	1.87e-92
45	530	14.4	1420 2 A32869		

ALIGNMENTS

RESULT 1
ENTRY A43090 #type complete
TITLE enteropeptidase (EC 3.4.21.9) precursor - bovine
ALTERNATE_NAMES enterokinase
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

ACCESSIONS
REFERENCE A43090; A46874; A61436
#authors Kitamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.
#journal Proc. Natl. Acad. Sci. U.S.A. (1994) 91:7588-7592
#title Enterokinase, the initiator of intestinal digestion, is a
mosaic protease composed of a distinctive assortment of
domains.

#cross-references MUID:94329561
#accession A43090
#status nucleic acid sequence not shown; translated from
GB/EMBL/DBJ

#molecule_type mRNA
#residues 1-1035 #label KIT
#cross-references GB:U09859; NID:9746410; PID:AA40026.1; PID:9746411
#experimental_source small intestine
#accession A46874

#authors Lavallo, E.R.; Rehembulla, A.; Racle, L.A.; DiBasilio, E.A.;
Ferent, C.; Grant, K.L.; Light, A.; McCoy, J.M.

#journal J. Biol. Chem. (1993) 268:23311-23317
#title Cloning and functional expression of a cDNA encoding the
catalytic subunit of bovine enterokinase.

#cross-references MUID:94043122
#accession A46874

#molecule_type mRNA
#residues 801-1035 #label LAV
#cross-references GB:U19663; NID:9416131; PID:AA16035.1; PID:9416132
#note parts of this sequence, including the amino end of the
mature protein, were confirmed by protein sequencing

#journal Light, A.; Janska, H.
#journal J. Protein Chem. (1991) 10:475-480
#title The amino-terminal sequence of the catalytic subunit of
bovine enterokinase.

#cross-references MUID:92189715
#accession A61436

#molecule_type protein
#residues 801-807, 'Y', 809-827 #label LIG
#comment The mechanism of association with the membrane of the intestinal
brush border is unknown. The sequence is compatible with
amino-terminal myristoylation of the heavy chain or with type II

AC V61239: (first entry)
 DT 06-JAN-1999
 DE cDNA sequence of prostate tumour clone.
 KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
 OS Homo sapiens.
 PN M09837093-A2.
 PD 27-AUG-1998.
 PE 25-FEB-1998: US03492.
 PR 09-FEB-1998: US-020956.
 PR 25-FEB-1997: US-806099.
 PR 01-AUG-1997: US-904804.
 PA (CORI-) CORIXA CORP.
 PI Dillon DC, Xu J;
 WP: 98-609886/51.
 DR Polypeptides comprising immunogenic portions of prostate proteins -
 PT used in a vaccine for the treatment of prostate cancer
 PS Claim 3; Page 97; 130p; English.
 CC The present sequence is a new DNA which encodes an immunogenic portion
 of a prostate tumour protein. The encoded immunogen, or the DNA itself,
 can be used as a vaccine for the treatment of prostate cancer. The DNA
 CC was identified by analysis of a substracted cDNA library obtained by
 CC substracting a prostate tumour cDNA expression library with a normal
 CC tissue cDNA library.
 CC Sequence 196 BP; 53 A; 58 C; 41 G; 44 T;
 SQ

Query Match 11.3%; Score 196; DB 1; Length 196;
 Best Local Similarity 100.0%; Pred. No. 8.3e-46;
 Matches 196; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	46	agcgcgcaagtcataatgaacatccagatccatcatctactcgaagcttgataac	105
	1	AGCGGCGCAGGCAATTCAGATTCAGATTCATTCATTCAGTCGTTGATAAC	60
QY	106	agcaagatcgagcttgactcaaggtcacaccacagctattgacctaactatgaaccat	165
	61	AGCAAGATGCGCTTTTAATCACTCAGGGTCACACACAGCTATTGGACCTTACATGAAAAACAT	120
DB	121	GGATTCACCAACGGGAAACCCCTATCCCGCACAGCCACCTGTGTCGCCACTGTCTACGAG	180
QY	166	gataccaacccggaaccctcctacccgcagacagcccaactgtgtgtcccaactgtaagag	225
	121	GGATTCACCAACGGGAAACCCCTATCCCGCACAGCCACCTGTGTCGCCACTGTCTACGAG	180
QY	226	gtgcataccggtccagt	241
	181	GTGCATCCGCGCTCACT	196
DB	181	GTGCATCCGCGCTCACT	196

RESULT 3
 V58638/c
 ID V58638 standard; cDNA; 195 BP.
 AC V58638;
 DT 08-DEC-1998 (first entry)
 DE Prostate tumour specific gene clone.
 KW Prostate tumour specific gene; human; prostate cancer; detection;
 KM Therapy; ss.
 OS Homo sapiens.
 PN M09837418-A2.
 PD 27-AUG-1998.
 PE 25-FEB-1998: US03690.
 PR 09-FEB-1998: US-904809.
 PR 25-FEB-1997: US-806596.
 PR 01-AUG-1997: US-904809.
 PA (CORI-) CORIXA CORP.
 PI Dillon DC, Xu J;
 WP: 98-480805/41.
 DR Novel human prostate specific tumour protein and fragments - useful
 PT for detecting and treating prostate cancers
 PS Claim 1; Page 109; 141p; English.
 CC This sequence represents a human prostate tumour specific gene, and can
 be used in the method of the invention. The method is for detecting
 CC prostate cancer comprises contacting a biological sample with an agent
 able to bind an immunogenic portion of a prostate protein (such as
 CC encoded by this sequence). An antibody which binds to an immunogenic

CC	portion of the prostate protein, and the method can be used to detect,
CC	monitor progression of, or treat prostate cancers. The antibody may
CC	also be conjugated to a therapeutic agent for use in therapy of prostate
CC	cancers.
SQ	Sequence 195 BP; 40 A; 51 C; 43 G; 60 T;
Query Match	11.2%; Score 194; DB 1; Length 195;
Best Local Similarity	99.5%; Pred. NO. 3e-45; 1; Indels 0; Gaps 0;
Matches 194; Conservative	0; Mismatches
QY	563 actctctcagaaggaagcctctggaacctgtgtgcacaaagacagctggaaacgaactacg 652
Db	195 ACTCTCTTCAGAGGAAGTCCTGGCACCCCTGTGTGCCAAGACACATCGGAACGAACCTACG 136
QY	653 ggcggcgcgccctgcagggagacatgggctctaagaatatttttaactctagccaaagatatg 712
Db	135 GCGCGGCGCGCCCTGCAGGGAGCATGGGCTTTACAAATAATTTTACTCTTACGCCAAGGAATG 76
QY	713 tggatgacagcggatccaccacagcttctatgaaactgaacacaagctgcggcgaatgctcgata 772
Db	75 TGGATGACAGCGGATCCACACAGCTTTATGAACCTGAACACAAAGTGCCGCAATGTCGATA 16
QY	773 tctataaaaaactgt 787
Db	15 TTTATATAAAACTGT 1

	RESULT	4	
ID	V61319/c		
AC	V61319 standard; cDNA; 195 BP.		
D	V61319; (first entry)		
DE	cDNA sequence of prostate tumour clone.		
KW	Prostate; cancer; tumour; vaccine; immunogen; clone; ss.		
OS	Homo sapiens.		
PN	M0837093-A2.		
PD	27-AUG-1998.		
PF	25-FEB-1998.		
PR	09-FEB-1998; US-020956.		
PR	25-FEB-1997; US-806099.		
PA	01-AUG-1997; US-904804.		
PI	(CORI-) CORIXA CORP.		
DR	Dillon DC, Xu J; MPI: 98-609886/51.		
PT	Polypeptides comprising immunogenic portions of prostate proteins - used in a vaccine for the treatment of prostate cancer		
PS	Claim 12; Page 102; 130pp; English.		
CC	The present sequence is a DNA which encodes an immunogenic portion of a prostate tumour protein. The encoded immunogen, or the DNA itself, can be used as a vaccine for the treatment of prostate cancer. The DNA was identified by analysis of a subtracted cDNA library obtained by subtracting a prostate tumour cDNA expression library with a normal tissue cDNA library.		
SQ	Sequence 195 BP;	40 A;	51 C; 43 G; 60 T;
	Query Match	11.2%;	Score 194; DB 1; Length 195;
	Best Local Similarity	99.5%;	Pred. No. 3e-45; Indels 0; Gaps 0.
	Matches 194; Conservative	0;	Mismatches 1;
OY	593 actcatctcaaggaggaagtcttgacaccctgtgtgcgaagaagcactggaacgaagaactacg	652	
DB	195 ACTCATCTCAGAGGAAGTCTTCGGACCCGTGTCGCCAAGACATCGAAGCAAGAACTACG	136	
OY	653 gggcgggcgagctctgcagggaacatgggacctataagaataatttttaaccttcagccaagaatc	712	
DB	135 GCCGGGCGGCGCCGACGGGACATGGGCATTAAATAAATTTTACTCTACCCAAGGAATGG	76	
OY	713 tggatgcagcggcgatccaccacagcttatgtaaactgaacacaagaatgccggcaatgtcgata	772	
DB	75 TGAGTGACACGGCGATCCACACAGCTTTATGAACTGACACAAGAAGTCCGGCATGTGTGATA	16	

785-1019	#product enteropeptidase light chain #status predicted
785-1014	#label LCN\
116,147,179,328,	#domain trypsin homology #label TRY\
335,388,440,470,	
503,534,630,682,	
706,725,848,887,	
909,949	#binding_site carbohydrate (Asn) (covalent) #status predicted\
772-896, 810-826,	
910-977, 941-956,	
967-995	
825,876,971	#disulfide_bonds #status predicted\
SUMMARY	#active_site His, Asp, Ser #status predicted
	#length 1019 #molecular_weight 112923 #checksum 7518
Query Match	24.2% Score 890; DB 1; Length 1019;
Best Local Similarity 36.1% Pred. No. 5,25e-179;	
Matches 137; Conservative 86; Mismatches 140; Indels 17; Gaps 16;	
Db 643 CKADHFOCKN-GECEVPLVNLCDGHLCEDESDSEADCVREFNGTNNGLVRFQSI-NH 700	
Oy 113 CNSGIEDESSGTCINPSNMCDGVSHPGCEDENRCLY-CP-NFI-LQVYSSQRRSWH 169	
Db 701 TACAENWTTQTSNDVQCLLGSG-NSSKRIFFSDG-PRVKLNTKAPDG-HLLLT-PSQO 756	
Oy 170 PVEDQDMNENYGRACRCDMGCKNNFYSSGIVDSSGTSFMKLTSAQNVDIKKTYSQA 229	
Db 757 CLDPLRLQCNHNSCKKKLAODITPRKITYGSAKGAPEWVYGLYGGRLLCASLVS 816	
Oy 230 CSSKRAYSLKRI--ACGVNLNS-SRGRITVGSEALPGAMPQVSLHYQNVHYCGSIIIT 286	
Db 817 SDMLVSAACHVYGNLE-PSKWTAILGLHMKSNLTSPQTVPLRIDELIVINPHYRRKKDN 875	
Oy 287 PEWITTAACHV-EKPLNPNMWHTEFAGI-LRQSFMEFGAQY-Q-VEKYISHPNYDSKTKNN 343	
Db 876 DIAMMHLEFKVNYDYIOPICLPEBNVFFPRGRNCIAGKGYVYOGTANILQEDAVL 935	
Oy 344 DIALMKLOKPLTFNDLVKPYCLPRPGMLDPEQLCWISGVAIEERKGTSEVLNAKKVLL 403	
Db 936 ISNERCOOQ-PEPYNITENMICAEGEGIDSCQDGSGLPCQENRPFAGVTSFGYK 994	
Oy 404 IETGRCSRYVYDLNLTIPAMICAGFLOGNVNDSCQDGSGLVYSKNNIMWLTIDTISWGS 463	
Db 995 CALPNRPGVYAVARSFTEWI 1014	
Oy 464 CAKAYRPGVYGVVAFIDWI 483	

RESULT 3

ENTRY A53663 #type complete

TITLE enteropeptidase (EC 3.4.21.9) precursor - pig

ALTERNATE_NAMES #enterokinase

ORGANISM #formal_name Sus scrofa domestica #common_name domestic pig

DATE 07-Oct-1994 #sequence_revision 03-Aug-1996 #text_change 18-Jun-1999

ACCESSIONS A53663

REFERENCE A53663

AUTHORS Matsushima, M.; Ichinose, M.; Yahagi, N.; Kakei, N.; Tsukada S.; Miki, K.; Kurokawa, K.; Tashiro, K.; Shiohara, K.; Shikamaishi, K.; Uneyama, H.; Inoue, H.; Takahashi, T.; Takahashi, K.

#journal J. Biol. Chem. (1994) 269:19976-19982

#title Structural characterization of porcine enteropeptidase.

#cross-references MUID:94327548

#accession A53663

#molecule_type mRNA

#residues 1-1034 #label MAT

#cross-references GB:D30799; NID:9505122; PIDN:BA06459.1; PID:9505123

#note parts of this sequence, including the amino ends of three chains isolated from the mature protein, were determined by protein sequencing

COMMENT The mechanism of association with the membrane of the intestinal brush border is unknown. The sequence is compatible with type II

		membrane attachment using a signal-anchor sequence (as annotated below) or with amino-terminal myristoylation of the heavy chain.
COMPLEX		Mature enteropeptidase is variously reported to contain two (heavy and light) or three (mHnI, heavy, and light) chains derived from a single precursor form; heavy and light chains are linked by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms involves further processing. See the human entry (PIR:A56318) for an alternative version of products.
FUNCTION	#description	cleaves activation peptide from trypsinogen to produce active trypsin
CLASSIFICATION	#pathway #superfamily #receptor ligand-binding repeat homology; scavenger receptor cysteine-rich domain homology; trypsin homology glycoprotein; hydrolase; serine proteinase; transmembrane protein; zymogen	intestinal digestive hydrolyase cascade enteropeptidase, C1r/C1s repeat homology; LDL receptor ligand-binding repeat homology; MAM homology;
KEYWORDS		
FEATURE		
22-38		#domain transmembrane #status predicted #label TM\
52-117		#product enteropeptidase mHnI chain #status predicted #label MCH\
118-799		#product enteropeptidase heavy chain #status predicted #label HCN\
199-236		#domain LDlL receptor ligand-binding repeat homology #label LDlL1\
357-519		#domain MAM homology #label MAM\
541-646		#domain C1r/C1s repeat homology #label C1R\
658-692		#domain LDlL receptor ligand-binding repeat homology #label LDlL2\
693-798		#domain scavenger_receptor cysteine-rich domain homology #status atypical #label SRCR\
800-1034		#product enteropeptidase lght chain #status predicted #label LCH\
800-1029		#domain trypsin homology #label TRY\
116,147,170,194,		
283,343,350,403,		
455,485,518,549,		
645,697,701,721,		
740,761,804,863,		
902,964		
787-911,825-841,		#binding_site carbohydrate (Asn) (covalent) #status predicted\
925-992,956-971,		
982-1010		#disulfide_bonds #status predicted\
840,891,986		#active_site His, Asp, Ser #status predicted
SUMMARY	#length 1034 #molecular_weight 114793 #checksum 9505	
Query Match	23.5% ; Score 862; DB 1; Length 1034;	
Best Local Similarity	35.0%; Pred. No. 3,55e-172;	
Matches 133; Conservative	89; Mismatches 141; Indels 17; Gaps 16;	
Dbb	658 CKEDNFOCEN-GECVLVNLICDGFSHCKDGSDAHACVFRLNTANNSGLVQFRIOST-WH 715	I : :: I : I : I : I : I : I : I : I : I : I : I :
Oy	113 CSNSGIECDSSGTCINFSNMCGDVSHCPGEDENRCVR-LYG-PWFI-QYLYSSQRKSMH 169	: : : : : : : : : : : : : : : :
Dbb	716 TACAEMNTTQTSDVCQLGLLGIG-NSSMPFES-SGGGFYVLTNPAP-NGSLILTF-ASEO 771	: : : : : : : : : : : : : : : :
Oy	170 PYCODDMNENTGRAACKDMQYKKNFYSOGIVDDSSGSTSEFMKLINSAGNVDIYYKYKXSDA 229	: : : : : : : : : : : : : : :
Dbb	772 CFEDSLILLQCNRHSCKRKVOAEVPFKITGVGDSREGAMPVVALLYYNGOLLGCASVIS 831	I : :: I : I : I : I : I : I : I : I : I : I : I :
Oy	230 CSSRAAYVLRKI--AGCVLN-ISSRSQSRITVGSAALPGAMPQVSLHYOVNHVCGSIIIT 286	: : : : : : : : : : : : : : : :
Dbb	832 RDLVSAAHCYGRNLE-PSKMAIIGLMTSNLSPOLVTLLIDELIYNPHNYRRKDS 890	: : : : : : : : : : : : : : : :
Oy	287 PEWTVTAACHV-EKPLNNPWHMFAFGAI-LROSFMFYGAGY-VERYVLSHPDYSKTKNN 343	:
Dbb	891 DIAMMHLEPKVNTDTYOPICLDPEENOVDPPRGISIACMGSVVIYQGSPADILEDAPVL 950	I : : : : : : : : : : : : : : : :
Oy	344 DIALMKIQKPLETDLVKPCFLRPNGMILOPEDLCISMGMGLETKGTSEVLAQAVAL 403	:

D	694	TGTAGCCCGGACCTACACCCCATGCTGTGTGCAACTGTGGGGTTTCAAGCTGTGATCTATATATG	753
OY	1098	aaagatgattctcaccacaattatgatctccagagccagaagaacaatgatctgcgtgat	1157
D	754	GGGCTACCTTCCCTTCGAGACCTTACTATGACGAAACAGCAANTGACATTGCTTGGT	813
OY	1158	gaagctgcagaaagccctctgaactttcaagacctggtgaacacagtgctgtcccaacc	1217
D	814	CCACCTCTAGTCTCCCTGCTCTACAGAAATACATCCAGCAAGTGTGTCTCCCTGCTGC	873
OY	1218	agacgatgatctgcagccagaaacagctctgcgtgatttcgggtgaggccacccagga	1277
D	874	GGGACAGAGCCCTGTGTGTGATGCGAAGTCTGTACTGTGTACCCGGCTGGGTAAACACAGT	933
OY	1278	gaaaagggaagaccctcagaatgctgtaacgctgcacaaggtgctcttcattgaacacagag	1337
D	934	CTATTGGCAACAGCATGTATGCTGCTCCAAAGAGCCCGGGTTCCTCATTAAGCAACGAGT	993
OY	1338	atgcacaagcagatatgcttatgacaacacctgatccacacgatgatctgtccgctct	1397
D	994	TTTGCACAGCCCTCGACTTTACGGGAATCAGATCAAGCCCAAGATGTTCTGTGCTGCTA	1053
OY	1398	ctcgcaaggggaagctgcatctctgcacaggtgtacaagtggagggcccttggt-----	1449
D	1054	TCTGAGAGGTGGCATTTGATGCTGCTCCAGGCGACAGTGGAGGCCCTTTGTGTGTGAAGA	1113
OY	1449	---cacttcgaagaacaatatctgtgtgctgataaggagatacaagctgagggttcgtgctg	1505
D	1114	CAGCATCTCTGGGACATCAAGGTGGGGGCTATGTGTCATTGTAAAGCTGGGGTACGGGCTG	1173
OY	1506	tgcgaagctctacagaccagaggtgtaeaggaaatgtatgtattcagagactgagatta	1565
D	1174	TGCTTTGGCCCCGAGAACCCAGAGATGTACACCAAATGCTACATTCGCGAGTGTGATCTT	1233
OY	1566	tcgacaatgtagggcgacagcgtaatacca	1594
D	1234	CAGGCCATTAAGACTCACTCCGAAGCCA	1262
RESULT	7		
V59136			
ID	V59136 standard; DNA; 980 BP.		
AC	V59136;		
D	07-JAN-1999 (first entry)		
DE	Nucleotide sequence of SP0031A, a homologue of HELA2.		
KM	serine protease; regulation; cell activity; viability; HELA2; ATC2;		
KW	BCOM; testis; fertility; suppressor; testicular germ cell cancer;		
KW	semloma; testis-specific expression; antitumour; sperm development;		
KW	infertility; human; chromosome 16p13.3; ss.		
OS	Homo sapiens.		
FH	Key	Location/Qualifiers	
FT	CDS	3..899	
FT		/tag-a	
FT		/product=SP0031A	
FT		/note="sequence contains one internal stop codon"	
PN	NC0836054-A1.		
PD	20-AUG-1998.		
PF	13-FEB-1998; AU0085.		
PR	18-NOV-1997; AU-000422.		
PR	13-FEB-1997; AU-005101.		
PA	(AMRA-) AMRAD OPERATIONS PVTY LTD.		
PI	Antalis TM, Hooper JD;		
DR	WPI: 98-480768/41.		
DR	P-PDSH: W77304.		
PT	New serine protease(s) and kinase involved in regulating cell		
PT	activity and viability - particularly the testis-specific protease		
PT	HELA2 used for modulation of fertility and as tumour suppressor		
PS	Example 15; Fig 20C; 167PP; English.		
CC	V59134-36 represent HELA2 homologues. The genes are found in a cluster		
CC	on chromosome 16p13.3. HELA2 was isolated from HeLa cells, and has		
CC	homology to serine proteases. The protein is involved in or associated		
CC	with regulation of cell activity and/or viability. Administration of		
CC	recombinant HELA2 (also called testisin) is used to increase fertility		

Query Match	6.9%	Score 120	DB 1	Length 980
Best Local Similarity	50.3%	Pred. No. 3.8e-24		
Matches 371	Conservative 0	Mismatches 340	Indels 27	Gaps 2
Sequence	980 BP	192 A	325 C	289 G
				174 T
QY	858	ctcaagccgccaagcagagattgttgggggagagagcgcccccgggggagcttggccctg	917	
Db	11	CCCCGATGCTGAACCCAAATGTTGGGGGGGAGACACGCGAGAGGGCGAGTGGCCCTG	70	
QY	918	gcaagtcacgcttgcacgtcccaagacgctccacgcttgcgagagctccatcatcccccga	977	
Db	71	GCAAGTCAGCATCCAGGCCCAACGAGACCACTTTCGGGGGGCAGCCTCATCGCGAGCA	130	
QY	978	gtgatactgacagcgcccactgcgtgagaaaaaccccttacaacatcgaatgtgac	1037	
Db	131	GTGGATCTTGACGGCTGGGCGACCTGCTCCGCAACACCTCTGAGAGCTCCCTTACCAAGT	190	
QY	1038	ggcatttcgggagatttttgagacaaattcttcatgttcatgtagccggaatacgaagtga	1097	
Db	191	CTGTCTGGGGGCAAGCAGCTAGTCACGCGGAGCACACGCTATGATGCCGGGTGAG	250	
QY	1098	aaaagtattcttcataccaattatcgaactccaagaacaaatgaacatttcgctgat	1157	
Db	251	GCAAGTGGAGAGCAACCCCTCTTACCAGGGCACGGCCTTCACGCGTACAGTGGCCTGT	310	
QY	1158	gaagctgcagaagacgctctgacttctcaagcaccatgtgaaacagatgtgtcgtcccaacc	1217	
Db	311	GGAGCTGGAAGGACACAGTGGCCTTCACCAATATCAATCCGCCGATGGCTCTGCTGACCC	370	
QY	1218	aggcatgatctgcagcagaagaacgctctgtcgaattccgggttgggggacacaggagga	1277	
Db	371	CTCGGTGATCTTTGAGAGCGGGCATGATACCTGCTGGGTCACTGGCTGGGGGACGCCAGTGA	430	
QY	1278	gaagggagaagacctagaagatgtcttgaagcgtgcgaa-----gttgccttcatttgaagc	1331	
Db	431	GGAAACACTCTCTGCCCGGAACCGGGATCTGCGAAGAACTCGCTGTGTCCTCATATGACAC	490	
QY	1332	acaagatgcacaacagacagatatgtctatgacaaactg-----at	1370	
Db	491	ACCCAAAGTGAACCTGCTCTACAGCAAAAGACACGAGTTTGCTTACCAACCAAAACAT	550	
QY	1371	caacacagccatgatctgttcgcggtcttcctcagggagaaagtcgattcttcacagggtga	1430	
Db	551	CAAGATATACATGCTGTGCGCGCTTCGAGAGAGGGCAAGAAAGATATGCTCGCAAGGCCA	610	
QY	1431	cagtggaagcgctctgtgtcaacttcgaagaacaaatacttctgtgtcgaatagggatacaag	1490	
Db	611	CTCGGGGGGGCCCCCTGGGTGTGCTCTGTGGGTGATGATGCTGGCTGACAGCGGGGGTATCAG	670	
QY	1491	ctggagcttcgtcgttgcgaagaacttaacagacagagatgttaccggaaatgttgaatgtat	1550	
Db	671	CTGGGGGTAGAGGCTTGCCCGGCAAGACCGCCAGAGTGTCTTACATCCGATGTCACCGCCCA	730	
QY	1551	caaggactggaattatcg	1568	
Db	731	CCACAACGTGATCATCG	748	

```

#molecule_type mRNA
##residues 1-638 #label SEI
##cross-references GB:M30287; NID:g205010; PID:AAA4163.1; PID:g205011
part of this sequence, including the amino ends of both
the heavy and light chains, was confirmed by protein
sequencing

REFERENCE
#authors S06851
#journal M.; Seldah, N.G.
#title Blochim. Biophys. Acta (1989) 999:103-110
#cross-references GB:M30287; NID:g205010; PID:AAA4163.1; PID:g205011
sequencing and development of a specific radioimmunoassay
#accession S06851
#molecule_type protein
#residues 20-45;391-413 #label PAO
REFERENCE
#authors Seldah, N.G.; Ladenheim, R.; Mbikey, M.; Hametlin, J.;
#journal Lutfalla, G.; Rougeon, R.; Lazure, C.; Chretien, M.
#title DNA Cell Biol. (1988) 8:563-574
#accession I53041
The cDNA structure of rat plasma kallikrein.
#status translated from:GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-638 #label RES
##cross-references GB:M58590; NID:g206721; PID:AAA42069.1; PID:g206722
COMMENT This protein, synthesized in the liver, circulates as a noncovalent
complex with high molecular weight (HMW) kininogen.
COMMENT The zymogen is activated by factor XIIa, which cleaves the molecule
into a light chain, which contains the active site, and a heavy
chain, which associates with HMW kininogen. These chains are
linked by one or more disulfide bonds.

GENETICS
#gene PK
CLASSIFICATION #superfamily coagulation factor XI; trypsin homology
KEYWORDS blood coagulation; duplication; fibrinolysis; glycoprotein;
hydrolyase; inflammation; liver; plasma; serine proteinase;
zymogen

FEATURE
1-19 #domain signal sequence #status predicted #label SIG\
20-390 #product plasma kallikrein heavy chain #status
experimental #label MAT1\
20-109 #domain apple repeat #label AP1\
110-199 #domain apple repeat #label AP2\
200-289 #domain apple repeat #label AP3\
291-380 #domain apple repeat #label AP4\
391-638 #product plasma kallikrein light chain #status
experimental #label MAT2\
391-621 #domain trypsin homology #label TRY\
21-104;47-77,51-57,
111-194,137-166,
141-147;201-284,
227-256;231-237,
292-375,318-347,
322-328,340-345,
383-503,419-435,
517-584,548-563,
574-602
#disulfide_bonds #status predicted\
127,215,308,453,
459,494
#binding_site carboxydrate (Asn) (covalent) #status
predicted\
396 #binding_site carboxydrate (Asn) (covalent) #status
experimental\
#active_site His, Asp, Ser #status predicted
#length 638 #molecular_weight 71273 #checksum 227
SUMMARY
434,483,578 #length 638 #score 740; DB 1; length 638:
Query Match 20.1%; Score 740; DB 1; length 638:
Best Local Similarity 44.5%; Pred. No. 1,42e-142:
Matches 105; Conservative 45; Mismatches 78; Indels 8; Gaps 6;

Db 389 ARIYGGTNSL-GEMPVSLQVLTVSQNMGGSGISGROMILTAHCFD-GIYPDPVNR 446
:::1::1::1::1::1::1::1::1::1::1::1::1::1::1::1::1::1::1::1::1
253 SRIVGG-ESALPGMHPQVSLHVNQV---HYVGGSTITPEMTYTAACVCEKPLNNPHNWT 308

```

Db	Accession	Entry	Title	Organism	Date	Accession	Entry	Title	Organism	Date
Db	447	IGGILILSETKNTKPPSSIKELIHOKKMSGSDVILITLOPLANTEROKRIPS	506							
Oy	309	AFAGILKQSNFTFGAGIO-VKXISHPNDSKTKKNDILMLKLPFLFNDLVKVCAPN	367							
Db	507	KADNTITTCWYTGWGYTKERGETONILQKATIPLYPBECKOXYR-DYVITKOMICAG	565							
Oy	368	PGMHLPEOLCISGMKATEKKEKSEVLNAKVLIIETROGNSRXYVDNLITPMICAG	427							
Db	566	YKEGIDACKGDSGGPLVCKHBSRMOLVGITTSGECCAKKEOPCYTTKVAETIDMI	621							
Oy	428	FLQGNVDSGCGDSGAPMTSKNNIMWLIDTSGSCAAYRPGYGVNMFTDWI	483							
RESULT	6									
ENTRY		JE0315	#type complete							
TITLE		low-density lipoprotein receptor-related protein - mouse								
ORGANISM		#formal_name Mus musculus #common_name mouse								
DATE		16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 13-Aug-1999								
ACCESSION		JE0315								
REFERENCES		JE0315								
#authors		Tomlita, Y.; Kim, D.; Magoori, K.; Fujino, T.; Yamamoto, T.T.								
#journal		J. Biochem. (1998) 124:784-789								
#title		A novel low-density lipoprotein receptor-related protein with type II membrane protein-like structure is abundant in heart.								
#accession		JE0315								
#status		preliminary								
##molecule_type		mRNA								
##residues		1-1113 #label TOM								
##cross-references		DDBJ:AB013874								
CLASSIFICATION		#superfamily trypsin homology								
FEATURE										
869-1097		#domain trypsin homology #label TRY								
SUMMARY		#length 1113 #molecular_weight 122984 #checksum 224								
Query Match		20.0%; Score 733; DB 2; Length 1113;								
Best Local Similarity		34.9%; Pred. No. 6, 93e-141;								
Matches 136; Conservative		70; Mismatches 158; Indels 26; Gaps 23;								
Db	723	CQDNLECANHE-CVPRDLWCQDGVDCSDSDSEWCQCVTLKNGNSSLLTVHKA-KH	779							
Oy	113	CSNGIGCDSSGCGINCINSNMCQDVSHCPGGEDENRCLVRL-VG-PNFIQIYVSSORKSMH	169							
Db	780	HYCADGGRFELSLQACKQMGLEPSTYKLI-IPQBEQO-WLILYNNWELNINSTIQLELLY	837							
Oy	170	PVCDDPNNENYGRACAKDMGKNNETYSOGIYDDSGSTSFMLNYS-AG-NVDIYKTY--	225							
Db	838	YRHSQSRSEISLSCSKODCGRRPARRMKRIILGRTSPRGMPQCSIQSEPSGHICG	897							
Oy	226	HSDDSCSKAYVSLRCL--ACGVNLNSRSRSLYVGSALPAGAMPQVSLHVG-NHVGCG	282							
Db	898	VLAKKKAVLTVAHCFESGEDADAVK-VFEGINNLDHSPCFMOTRP-VKTIILHPRISAV	955							
Oy	283	SIIPSEWIVAAHCE-KPLNNPMHTATA-GILRQSFMYGAGYQVEKVISHPYDSEKT	340							
Db	956	VDYVISVETLSDDINERSYRVPCLPSPSEYLEPDYCYITTMGHMGNK-MPFK-LOEGE	1013							
Oy	341	KNDIALMKIQLKPTFNDLVKPLCLRPGMMLQPEOLCISMGATEEKGTSEVLNAAK	400							
Db	1014	VRIIPBQOS-Y-FDMKTTINMIGAGYESTGVSCMSDGGPLVCEPBGQWTLGTLT	1071							
Oy	401	VLIETORNSRXYVD-NLITPMICAGFLQGVNDSGCGSGPLVTSK-NNIMWLIDGT	458							
Db	1072	SMGVSCTSKVLGPGVSNVSYFGWLEROI	1101							
Oy	459	SMWGGC-AKAYRPGYGVNMVFTDWI	487							
RESULT	7									
ENTRY		KOMSP	#type complete							
TITLE		plasma kallikrein (EC 3.4.21.34) precursor - mouse								


```

388-625      #product coagulation factor xia light chain #status
388-618      #experimental #label LCH\
20-103,514-581, #domain trypsin homology #label TRY\
571-599      #disulfide_bonds #status predicted\
29           #disulfide_bonds interchain #status experimental\
46-76,50-56, 110-193,136-165,
140-146,200-283, 226-255,230-236,
291-374,317-346, 321-327,380-500,
416-432,545-560, 90,126,353,450
339          #disulfide_bonds #status experimental\
387-388      #binding_site carbohydrate (asn) (covalent) #status
431,480,575  #disulfide_bonds interchain #status predicted\
491          #cleavage_site Arg-Tile (coagulation factor xia) #status
          #active_site His, Asp, Ser #status predicted\
          #binding_site carbohydrate (asn) (covalent) #status
          experimental
SUMMARY      #length 625 #molecular_weight 70109 #checksum 9314
Query Match  18.9% Score 694; DB 1; Length 625;
Best Local Similarity 42.6%; Pred. No. 1,71e-131;
Matches 100; Conservative 43; Mismatches 86; Indels 6; Gaps 5;

Db 386 PRVCGTASVNGEPMQVTLTFTSPTRHLCGGSTIGQWLTAAHCY-GVESKILRV 444
QY 253 SRIVGGSALEPGAMPVQVSLHVN-V-HVCGGSITPEWITVAHCVEKPLNPMHTA 309
Db 445 YSGILNSEIKEDPSFGVCEIITHDQYKMAESGDIALLETTVNTDQRPCLPSK 504
QY 310 FAGILROSFMFYGCY-QVEKIVISHPNDSKTKNNIDALMLQKPLFNDLVKPCLEPN 368
Db 505 GDRNVITDCWVTGMYRKLRDKIONTLOKAKIPLVNEECQKRR-CHKTHKIKACY 563
QY 369 GMLQPOLCISGMCATEEKGTSEVLNAKVLLETQRCNSRYVDNLITPAMICAGF 428
Db 564 REGGDKCKDGGSGPLSCKHNEVHVICITSMGSCAQREPPGYTVNVEVDWI 618
QY 429 LQGNVDSGCGSGPLVTSKNNIMWLTGDTSMGSCAKAYRPGVNVVETDWT 483

RESULT 9
ENTRY      S00845 #type complete
TITLE      hepsin (EC 3.4.21.-) - human
ORGANISM   #formal_name Homo sapiens #common_name man
DATE       31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change
18-Jun-1999
ACCESSIONS S00845
REFERENCE   S00845
AUTHORS    Leytus, S.P.; Loeb, K.R.; Hagen, F.S.; Kurachi, K.; Davie, E.W.
#journal   Biochemistry (1988) 27:1067-1074
#title     A novel trypsin-like serine protease (hepsin) with a putative transmembrane domain expressed by human liver and hepatoma cells.
#cross-references MIMD:88209431
#accession S00845
#molecule_type mRNA
#residues 1-417 #label LEY
GENETICS   ##cross-references EMBL:X07732; NID:g32063; PIDN:CAA30558.1; PID:g32064
#gene      GDB:HPN; TMRPSS1; hepsin
#map_position 19q11-19q13.2
CLASSIFICATION #superfamily hepsin; trypsin homology
KEYWORDS     hydrolase; liver; serine proteinase; transmembrane protein
FEATURE      23-45 #domain transmembrane #status predicted #label TMN\
163-400      #domain trypsin homology #label TRY\
188-204,291-359,

```

```

322-338,349-381 #disulfide_bonds #status predicted\
203,257,353      #active_site His, Asp, Ser #status predicted
SUMMARY          #length 417 #molecular_weight 45011 #checksum 5762
Query Match      18.5% Score 680; DB 1; Length 417;
Best Local Similarity 33.4%; Pred. No. 3.9e-128;
Matches 120; Conservative 84; Mismatches 135; Indels 20; Gaps 15;

Db 54 VQVSADARLWVFDKTEGTWRLCCSRNARVAGLSCEMFLRALTSHSELDVPTAGANG 113
QY 149 VRLGPNFILOVSSQKSMHPVQODDMENYGRACDKCKNNFSSQ-GI-VDDS-G 205
Db 114 TSFEFCDEGRPLPHTORLELVISVDCPRGRFLAICODGGR-KLPVD-RIVGPRSL 171
QY 206 STSFKMLNTSA-GNVDYKRYHSD-ACSSKAVSLRCAICACVNNSSRSHRISVG-ESAL 262
Db 172 -GRPMQVSLRVDGAHLCCGSLSGDWLTAAHCFPEENRRLSWRRPAGVAASP-HG 229
QY 263 PGAMPQVSLHVNQVHCGGSITPEWITVAHCVEKPLNPMHTAFAGILROSFMFYG 322
Db 230 LQGVQAVVYHGTLFPRDPNPEENSENDIALVHLSPLPLEYIOPVCLPAAAGALVDGK 289
QY 323 AGVQEKYISHPNY----DSKTK-N-NDIALMLQKPLFNDLVKPCLEPNPMHTA 376
Db 290 ICTYTGNGNTQYVQQAQVLOEARVPIISNDVCGADFYGNQIRPKMFCAGYRPGGIDAC 349
QY 377 LQWISGMCATEEKGTSEVLNAKVLLETQRCNSRYVDNLITPAMICAGFLOGNDVSC 436
Db 350 QGDSGGPRVCDISIKRPRMLCGIVSGTGCALAQKRGVTKYSDREMFQALXHS 408
QY 437 QGDSGGPLY---SKNNIMWLTGDTSMGSCAKAYRPGVNVVETDWTIRQMRADG 491

RESULT 10
ENTRY      S33777 #type complete
TITLE      hepsin (EC 3.4.21.-) - rat
ORGANISM   #formal_name Rattus norvegicus #common_name Norway rat
DATE       06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change
18-Jun-1999
ACCESSIONS S33777; S32013
REFERENCE   S33777
AUTHORS    Farley, D.; Raymond, F.; Nick, H.
#journal   Blochim. Biophys. Acta (1993) 1173:350-352
#title     Cloning and sequence analysis of rat hepsin, a cell surface serine proteinase.
#cross-references MIMD:93305733
#accession S33777
#status    preliminary
#molecule_type mRNA
#residues 1-416 #label PAR
CLASSIFICATION #superfamily hepsin; trypsin homology
KEYWORDS     hydrolase; liver; serine proteinase; transmembrane protein
FEATURE      22-44 #domain transmembrane #status predicted #label TMN\
162-399      #domain trypsin homology #label TRY\
187-203,290-358, #disulfide_bonds #status predicted\
321-337,348-380  #active_site His, Asp, Ser #status predicted
202,256,352      #length 416 #molecular_weight 44926 #checksum 5191
SUMMARY      #length 416 #molecular_weight 44926 #checksum 5191
Query Match      17.8% Score 652; DB 1; Length 416;
Best Local Similarity 40.6%; Pred. No. 2.0e-121;
Matches 101; Conservative 52; Mismatches 83; Indels 13; Gaps 9;

Db 161 RIVGGODSSL-GRPMQVSLRVDGTHLCCGSLSGDWLTAAHCFPEENRRLSWRRVAG 219
QY 254 RIVGG-ESALPGAMPQVSLHVNQVHCGGSITPEWITVAHCVEKPLNPMHTAFAG 312
Db 220 AVARTSP-HAVQGVQAVYHGGTLFPRDPIDNSNDIALVHLSPLPLEYIOPVCLP 278
QY 313 ILROSFMEYAGVQYKISHPNY----DSKTK-N-NDIALMLQKPLFNDLVKPCLEP 366

```

The invention relates to 86 novel genes and their fragments (nucleic acid sequences: X04311-X04410; amino acid sequences W78126-W78225) which are useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. Also, pathological conditions can be diagnosed by determining the amount of the new polypeptides in a sample or by determining the presence of mutations in the new polynucleotides. Specific uses are described for each of the 86 polynucleotides, based on which tissues they are most highly expressed in (see X04311 for described uses).

Sequence 1008 BP; 253 A; 278 C; 265 G; 208 T;

Query Match 5.8%; Score 100.4; DB 1; Length 1008;
Best Local Similarity 53.6%; Pred. No. 1,2e-18;
Matches 253; Conservative 1; Mismatches 212; Indels 6; Gaps 2;

QY 1143 tgacattgagctgctgagctgcaagaagctctgactttcaagactatgaaacagt 1202
DB 1 TGACATCGCCCTCATGAGCTGCAAGTCCACATCTTCTCAGGCACACTCAGGCCCAT 60
QY 1203 gttctgcccacccagcagctgactgctcagcagaaagctctgctgattccgggtg 1262
DB 61 CTGCTGCCCCCTTTTGATGATGAGAGCTCACTCCAGCCACTCTGATCATGTGATG 120
QY 1263 gggggccaccgagagaga---agggaagaccctcagaagtgctgaagcgtgccaagtgtct 1319
DB 121 GGGCTTTACAGACAGAAATGAGAGGAAATGCTGACATATGCTGTCAGGCGTCAGTCA 180
QY 1320 tctcaattgagacagagatgcaaacagcagatgctatgacaaacctgtacacagc 1379
DB 181 GGTCAATTGACAGACAGCGGMAATGACAGATGCTGACAGGCGGAGTCAACCGAGAA 240
QY 1380 catgactgtgcccggctctctcagaggaagcagctctctcagaggtgacagtgaag 1439
DB 241 GATGATGTGTGACAGGCTATCCCGAAGGGGTGTGACACTCTCCAGGCTGACATGTG 300
QY 1440 gctctgctgactctgagagaacaatatctgtgtgctgataagggtatacaagctgggttc 1499
DB 301 GCCCCTG---ATGACCAATCTGACACGTGCAATGTGTGGCATCTGTTAGCTGGGCTA 357
QY 1500 tggctgtgccaagactcacaagcagagagtgctacaggaatgtgattatcaagactg 1559
DB 358 TGGCTGGGGGGCCGACGCCACCCAGGAGTATACACCAAGGTCTCACCTATCTCACTG 417
QY 1560 gatttatcgacaatatgaggcagagcagctaatccaatgltcttcgctctg 1611
DB 418 GATCTACAAATGTCTGAGGCTGAGCTGTAATGCTGCTGCCCTTTGCACTG 469

RESULT 13
062299
ID 062299 standard; cDNA; 1440 BP.
AC 062299;
DT 27-MAY-1994 (first entry)
DE Factor VIII.
KM Truncated tissue factor; tTF; factor VIIa; FVIIa; activator;
KW bleeding disorder; haemophilia; liver cirrhosis; coagulation;
KW transmembrane domain; extracellular domain; soluble; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 35..1436
FT /*tag- a
FT /product- factor_VIII
MO933074-A.
PD 25-NOV-1993.
PE 12-MAY-1993; U04493.
PR 13-MAY-1992; US-882202.
PR 19-FEB-1993; US-021615.
PA (OKLA-) OKLAHOMA MED RES FOUND.
PI Comp PC, Morrissey JH.
DR WPI: 93-386218/48.
DR P-PSDB: R52562.
PT Compn. of truncated tissue factor and factor-VIIa or activator

PT of VII - useful for treating bleeding disorders, e.g. haemophilia
or liver cirrhosis
PS Disclosure; Page 26-28; 43pp; English.
CC A compn. for treatment of patients with prolonged or excessive
CC bleeding disorders comprises truncated tissue factor (tTF) (052439)
CC and a substance effective to produce a plasma level of factor VIIa
CC which in combination with tTF will control or stop the excessive
CC bleeding. The substance may be factor VIIa itself or an activator
CC which promotes the conversion of endogenous factor VII to factor VIIa.
SQ Sequence 1440 BP; 288 A; 434 C; 464 G; 254 T;

Query Match 5.5%; Score 95.4; DB 1; Length 1440;
Best Local Similarity 47.9%; Pred. No. 3.6e-17;
Matches 353; Conservative 0; Mismatches 366; Indels 18; Gaps 2;

QY 865 cgcagagagcagattgttgaggcggagagcgcgtccggggccctggagcagcgtc 924
DB 660 CCCCAGGCCGAAATTTGGGGGCAAGGTGTGCCCAAGGGAGTGTCCATGTGCAAGTTC 719
QY 925 agcctgacgtccagaaagctcagctgtcgagagctccatcacaccctcagagtgatc 984
DB 720 CTGTTGTGTGTAATGAGAGTCAAGTGTGTGGGGGAGACCTGATCAACACATCTGGGTG 779
QY 985 gtgacagcgcgcacatgctgtgaaaaacctcttaacaatccatgcatggaagcatt 1044
DB 780 GTCTCCCGCCGCCCATCTTTCGACAAATCAAGAACTGAGAACTGATCCGGGCTGCTG 839
QY 1045 gcgaggattttggaacaattctcatgctctatgagggcggatcacaaagtagaagaatg 1104
DB 840 GCGAGAGACAGACTCAGACGACAGCGGGATGACAGACCGCGGCTGCGCCAGCTG 899
QY 1105 attctcatcacaatatatgactccaagaccagaagaatgacattgctgtatgaaagctg 1164
DB 900 ATCATGCCAGCAACGATAGTCCGGGACACCAACAGACAGATTCGGGCTGCCCTG 959
QY 1165 cagaagcctctgacttcaagacctagtgaaacagtgctgtcccaa-----c 1215
DB 960 CACACGCCCTGTGCTCCATCACTGACATGTTGTGTCCTCTGCTGCCCGAAGCAGCTTC 1019
QY 1216 ccaggcattgtgtctgacagaaacagctctgtgtgattccgggtggggggccacagag 1275
DB 1020 TCTGAGAGAGAGCTGCGCTTCGTCGCTTCTCATTTGCTCAGCGGCTGGGCCAGCTG 1079
QY 1276 gagaagggagaaacctcagaagtgctgaaagctgccaaggtctctcatatgagacag 1335
DB 1080 GACCGTGGCCGACAGGCGCCCTGAGCTCATGTGCTCAAGCTGCCCGGCTGATGACCCAG 1139
QY 1336 agatgc-----aacagcagatgctctatgacaacctgatacaacacagcatgac 1386
DB 1140 GACTGCTGACAGATCAGCAGAGGTGGAGACTCCCAATATCAGAGATCATATGTTTC 1199
QY 1387 tggcgggtctctctgacggggaagctgacttcttcagagtgaaagtgaggcctctg 1446
DB 1200 TGTCCCGCTACTCTGGATGACAGAGACTCTCTGCAAGGGGAGCACTGAGGCCCATAT 1259
QY 1447 gtccattcgaagaacaatatctgtgtgtgctgataagggtatacaagctgggtgtctg 1506
DB 1260 GCCACCACTAAGCGGGGACAGGTGTACTGACAGGGGCAATCTGACGTGGGCCAGGGCTGC 1319
QY 1507 gccaaagctcacaagacagagtgatcaggaatgtgattatcaagagctgattat 1566
DB 1320 GCAACCGTGGGCCACTTTGGGGGTGTACACAGGCTTCCAGATACATCAGATGAGTGC 1379
QY 1567 cgacaatatgaggcaga 1583
DB 1380 AAGCTCATGCGCTCAGA 1396

RESULT 14
V40389
ID V40389 standard; cDNA; 1440 BP.
AC V40389;
PT

```

#accession A61545
#molecule_type protein
#residues 1-33:34-117 #label SCH
REFERENCE S17527
#author Schaller, J.; Straub, C.; Kaempfer, U.; Rickli, E.E.
#journal Protein Seq. Data Anal. (1991) 4:69-74
#title Complete amino acid sequence of equine miniplasminogen.
#cross-references MIMD:92052077
#accession S17527
#molecule_type protein
#residues 118-455 #label SC2
CLASSIFICATION #superfamily plasmin; kringle homology; plasminogen-related
#protein precursor homology; trypsin homology
#fibrinolysis; glycoprotein; hydrolase; kringle; plasma;
#serine proteinase; zymogen.
KEYWORDS
FEATURE 1-33,34-117,
118-455
1-33
#product plasminogen (fragments) #status experimental
#label PRO
#domain activation peptide (fragment) #status
experimental #label APV
34-117,118-225.
#product plasmin (fragments) #status experimental #label
MAR
37-114
118-455
#domain kringle homology #label KR4\
#product miniplasminogen #status experimental #label
MIN\
126-205
226-455
#domain kringle homology #label KR5\
#domain plasmin chain B #status experimental #label BCH\
226-448
267,310,405
#domain trypsin homology #label TRY\
#active-site His, Asp, Ser #status predicted
SUMMARY #length 455 #checksum 1118

Query Match 16.8%; Score 617; DB 2; Length 455;
Best Local Similarity 38.2%; Pred. No. 4,65e-113;
Matches 97; Conservative 66; Mismatches 79; Indels 12; Gaps 12;

Db 210 FDCGPKVPEKSGRIVGCGVAIAHSWPMQISRTFRGHCGLTSPPEVLTAAHCL 269
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 241 IACG-VNLNSR-QSRIVGESALPGAMPQVSLHYQ-NVHVCSSITPEMIVTAHCY 297
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 270 ERS-SRSTYKVVLTG-HHEIRLAQAQID-V-SKFLER-SRA-DIALIKSSPAIT 323
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 298 EKPLNNMWHMTAFAGILROSEFMFGAGYQVEKISHPNYDSTKRNDAKLOKPLTFN 357
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 324 QNVIRALLPRADYVVAWAEFVNGETODSSNAG-VLKAQCPVIEKNKCN-RYEYLN 381
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 358 DLVAPVCLPMPGMLLOPEOLCWMIGWTEKGTSEVLNAKVLLETORCNSRYV-D 416
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 382 GRVSTELCAGHYGVDSGCGSGPLVCFEKDKYITLOGVTSMGLCARPNKPGVYRV 441
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 417 NLIRPAICAGFLOGNDSCGDSGGLPVYSKNNIMWILIGTSMGSCAKYRGGVGNV 476
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 442 SSFIMWIERIMQSN 455
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 477 MVEFDWIRQMRAD 490
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 13
ENTRY PLPG #type fragment
TITLE plasmin (EC 3.4.21.7) precursor - p1g (fragment)
ALTERNATE_NAMES miniplasminogen
CONTAINS miniplasminogen
ORGANISM #formal_name Sus scrofa domestica #common_name domestic pig
DATE 07-Jul-1990 #sequence_revision 01-Nov-1996 #text_change
18-Jul-1997
ACCESSIONS S03733; A25834
REFERENCE S03733
#authors Schaller, J.; Marti, T.; Roesseler, S.J.; Kaempfer, U.;
#journal Rickli, E.E. (1987) 1:91-102
#title Amino acid sequence of the heavy chain of porcine plasmin.
#comment Comparison of the carbohydrate attachment sites with the

```

```

#accession S03733
#molecule_type protein
#residues 1-560 #label SCH
REFERENCE S03735
#author Brunscholz, R.A.; Lerch, P.G.; Schaller, J.; Rickli, E.E.;
#journal Lergler, W.; Manneberg, M.; Gillesen, D.
#title Eur. J. Biochem. (1981) 114:465-470
#comment Comparison of the primary structure of the N-terminal CNBR
#fragments of human, bovine and porcine plasminogen.
#cross-references MIMD:81212097
#accession S03737
#molecule_type protein
#residues 1-57 #label BRU
REFERENCE A25834
#authors Marti, T.; Schaller, J.; Rickli, E.E.
#journal Eur. J. Biochem. (1985) 149:279-285
#title Determination of the complete amino-acid sequence of porcine
#cross-references MIMD:85203907
#accession A25834
#molecule_type protein
#residues 450-790 #label MAR
FUNCTION #description
#dissolves the fibrin of blood clots; acts as a proteolytic
#factor in a variety of processes including embryonic
#development, tissue remodeling and tumor invasion; in
#ovulation it weakens the walls of the graafian follicle;
#also activates the urokinase-type plasminogen activator
#fibrinolysis
#superfamily plasmin; kringle homology; plasminogen-related
#protein precursor homology; trypsin homology
#fibrinolysis; glycoprotein; hydrolase; kidney; kringle;
#plasma; serine proteinase
KEYWORDS
FEATURE 1-790
1-77
78-560
84-162
166-243
256-333
358-435
450-790
461-540
561-790
561-783
30-54,34-42,84-162,
105-145,133-157,
166-243,169-297,
187-226,215-238,
256-333,277-316,
305-328,358-435,
379-418,407-430,
461-540,482-523,
511-535,547-665,
557-565,587-603,
679-746,709-725,
736-764
602,645,740
SUMMARY #length 790 #checksum 3770

Query Match 16.7%; Score 613; DB 1; Length 790;
Best Local Similarity 39.5%; Pred. No. 4,18e-112;
Matches 105; Conservative 52; Mismatches 94; Indels 15; Gaps 14;

Db 533 DYCVPOCVT-SSFDGKPKVPEKRCARVVGCVSIPLHSRPMQISRTFRGHCGLT 591
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 228 DACSSKAVVSLRQICNG-VNLNSR-QSRIVGESALPGAMPQVSLHYQ-NVHVCSSI 284
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

```


QY 352 actaagaagcaactgtgcatcaacttgaccctggggacccctcctcgtggtgagctgcgtg 411
Db 241 actaagaagcaactgtgcatcaacttgaccctggggacccctcctcgtggtgagctgcgtg 300
QY 412 ggcgtgctcactcctggaagttcattgagcagcaagtgtcccaactcctggtatgagtgc 471
Db 301 ggcgtgctcactcctggaagttcattgagcagcaagtgtcccaactcctggtatgagtgc 360
QY 472 gaccctcaggtactcctcacaacccctcactggtgtgtagtgccgtgacacatgcccc 531
Db 361 gaccctcaggtactcctcacaacccctcactggtgtgtagtgccgtgacacatgcccc 420
QY 532 ggcggggagagcaggaatcgtgtgttcgcctctacagacccaactcctcactcaggtg 591
Db 421 ggcggggagagcaggaatcgtgtgttcgcctctacagacccaactcctcactcaggtg 480
QY 592 tactcatctcagaggaagtccctggaacccctgtgtgcagaagaactggaacagaaactac 651
Db 481 tactcatctcagaggaagtccctggaacccctgtgtgcagaagaactggaacagaaactac 540
QY 652 gggcgggcgccctcagaggaatggtgtatagaataatttctcctagcagaagata 711
Db 541 gggcgggcgccctcagaggaatggtgtatagaataatttctcctagcagaagata 600
QY 712 gtcgatgacagcgatccaccagcttattgaaactgacaacaagtgcgcgcaatgtcgtat 771
Db 601 gtcgatgacagcgatccaccagcttattgaaactgacaacaagtgcgcgcaatgtcgtat 660
QY 772 atctataaaaaactgtatcccaagtgtcctgtctctcaaaaagcagtggttctcttcgc 831
Db 661 atctataaaaaactgtatcccaagtgtcctgtctctcaaaaagcagtggttctcttcgc 720
QY 832 tgtatagccgcgggggtgcaacttgaaactcaagccgcagagcagagattgtggggcgag 891
Db 721 tgtatagccgcgggggtgcaacttgaaactcaagccgcagagcagagattgtggggcgag 780
QY 892 agcgcgctcccgggggcgcttgccctgagcggtcagcgtcagcagacgtcccaagctg 951
Db 781 agcgcgctcccgggggcgcttgccctgagcggtcagcgtcagcagacgtcccaagctg 840
QY 952 tgcggagagctccatcatcaaccgccagtgatcgtgacagcgcccaactgcgtgtgaaaaa 1011
Db 841 tgcggagagctccatcatcaaccgccagtgatcgtgacagcgcccaactgcgtgtgaaaaa 900
QY 1012 ccccttaacaatcatgtgcatgttgagcagcattgcgggggttttgagaaacaattctcatg 1071
Db 901 ccccttaacaatcatgtgcatgttgagcagcattgcgggggttttgagaaacaattctcatg 960
QY 1072 tctcatgagcgcgatacccaagtagaaaaagtgttctcatcccaattatgactccaag 1131
Db 961 tctcatgagcgcgatacccaagtagaaaaagtgttctcatcccaattatgactccaag 1020
QY 1132 accaagaacaatgacatgtgcgtgtgacgtcagagcctcgtacttccaacgacctta 1191
Db 1021 accaagaacaatgacatgtgcgtgtgacgtcagagcctcgtacttccaacgacctta 1080
QY 1192 gtgaaaccagatgtgtcgtcccaaccgggcatgtgtgtagccagagaagcgtctgtg 1251
Db 1081 gtgaaaccagatgtgtcgtcccaaccgggcatgtgtgtagccagagaagcgtctgtg 1140
QY 1252 atttcggggtggggggcgacagaggaagaaggagagaccccaagagtcgtaagcgcgc 1311
Db 1141 atttcggggtggggggcgacagaggaagaaggagagaccccaagagtcgtaagcgcgc 1200
QY 1312 aaggtgtcttcattcgagacacagagatgcaacagcagatatgtctatgcaaacctgac 1371
Db 1201 aaggtgtcttcattcgagacacagagatgcaacagcagatatgtctatgcaaacctgac 1280
QY 1372 aacacagcagatgtgtgtgtgcgggtcttcctgacaggggaagtcgtatttctgcaagggtgac 1431
Db 1261 aacacagcagatgtgtgtgtgcgggtcttcctgacaggggaagtcgtatttctgcaagggtgac 1320

QY 1432 agtggaggcctctgtcacttcgaagaacaatatctgtgtgctgatataggatcaaacg 1491
Db 1321 agtggaggcctctgtcacttcgaagaacaatatctgtgtgctgatataggatcaaacg 1380
QY 1492 tgggtgtctgtcgtgtgccaagcttaacagccagaggtgtaaggggaattgtgagtattc 1551
Db 1381 tgggtgtctgtcgtgtgccaagcttaacagccagaggtgtaaggggaattgtgagtattc 1440
QY 1552 acggactgatttatcgacaatatgaggcagacgacctaa 1590
Db 1441 acggactgatttatcgacaatatgaggcagacgacctaa 1479
RESULT 11
US-09-342-749-1
Sequence 1, Application US/09342749
GENERAL INFORMATION:
APPLICANT: Wong, Alexander K.C.
APPLICANT: Tavtligian, Sean V.
APPLICANT: Teng, David H.-F.
APPLICANT: Myriad Genetics, Inc.
TITLE OF INVENTION: TMRPSS2 Is a Tumor Suppressor
FILE REFERENCE: 2318-202
CURRENT APPLICATION NUMBER: US/09/342,749
CURRENT FILING DATE: 1999-06-29
EARLIER APPLICATION NUMBER: US 60/091,044
EARLIER FILING DATE: 1998-06-29
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 1479
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1476)
FEATURE:
NAME/KEY: conflict
LOCATION: (724)
OTHER INFORMATION: Listed as T in GenBank Accession No. U75329
FEATURE:
NAME/KEY: conflict
LOCATION: (1347)
OTHER INFORMATION: Listed as C in GenBank Accession No. U75329
FEATURE:
NAME/KEY: conflict
LOCATION: (1466)
OTHER INFORMATION: Listed as A in GenBank Accession No. U75329
FEATURE:
NAME/KEY: conflict
LOCATION: (1471)
OTHER INFORMATION: Listed as A in GenBank Accession No. U75329.
FEATURE:
NAME/KEY: allele
LOCATION: (478)
OTHER INFORMATION: This base can be G or A with G being the more
OTHER INFORMATION: common allele. The codon will change from Val to
OTHER INFORMATION: Met.
FEATURE:
NAME/KEY: allele
LOCATION: (777)
OTHER INFORMATION: This base can be C or T with C being the more
OTHER INFORMATION: common allele. The codon is unaffected with both
OTHER INFORMATION: alleles encoding Gly.
FEATURE:
NAME/KEY: allele
LOCATION: (768)
OTHER INFORMATION: This base can be C or T with C being the more
OTHER INFORMATION: common allele. This is a silent polymorphism.

FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Billings, Lucy J.
 REGISTRATION NUMBER: 36,749
 REFERENCE/DOCKET NUMBER: PF-0227 US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-855-0555
 TELEFAX: 415-845-4166
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1077 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: SCORNOT01
 CLONE: 556016
 US-08-807-151-2

Query Match 57.1%; Score 992.4; DB 21; Length 1077;
 Best Local Similarity 99.8%; Pred. No. 3,6e-281;
 Matches 993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 721 agcgagatccagcgttattgaactgaacacagtgccggcaatgctgatatctataa 780
 DB 83 AGCGAGTCCACGAGCTTTATGAACATGACCAAGTCCGGCAATGCTGATCTATATA 142
 QY 781 aaactgacacagtgatgctgtctctcaaaagcagtggtttcttaactgtatagcc 840
 DB 143 AAAGTGCACACAGTGTGCTGCTCTTCAAAAGCAGTGTCTTCACTGATATAGCC 202
 QY 841 tgggggggcaacttgaaactaaagccgagagagagatctggggcgagagagcgctc 900
 DB 203 TGGGGGGTCAACTGAACTCAAGCCGACAGAGATGCTGGGGCGAGAGCGGCTC 262
 QY 901 cgggggggcttgccctggcaggtcagctcagctcagagacgtccagctgctggagagc 960
 DB 263 CCGGGGGGCTGCGCTGGCGAGTCAAGCTCAGCTCCAGAACTCCAGTGGCGGAGGC 322
 QY 961 tccatatacccccgcgagtgatgctgagacgcccgcctgctgggaaaaacctctaac 1020
 DB 323 TCCATATACCCCCCGAGTGTGATGCTGACAGCCGCCACTGCGGAAAAACCTTTAAC 382
 QY 1021 aatccatgagcttgagcagcagcttgcgggattttgagacaatcttccatctatgga 1080
 DB 383 AATCCATGCGATGAGAGCGATTTGGGGGATTTTGAGAAATCTTTCACTGCTATAGA 442
 QY 1081 gccggaataccaagtagaataatgattctcattcaataatactgaactcaagaacgaac 1140
 DB 443 GCCGATACCAAGTAAAGATGATTTCATCAAAATATGACTCCAGACCAAGAAC 502
 QY 1141 aatgacattgctgtatgaaagctgcagaaagccctctgaacttcaagaccctgagaaacca 1200
 DB 503 AATGACATTGCGTGTGATGAAGCTGACGAAGCCCTGACTTTCAACGACCTAGTGAACCA 562
 QY 1201 gctgctgcgcccaacccagcagatgctgcagccagcaagctctgctgattccggg 1260
 DB 563 GCTGCTGCTGCCCAACCCAGGATGATGCTGACCAAGACCTGCTGATTTCCGGG 622
 QY 1261 tgggggggcaacgagagaaaggaagacctcagaagtgctgaaagctgccaagtgctt 1320
 DB 623 TGGGGGGGCCACCCAGGATGATGCTGACCAAGACCTGCTGATTTCCGGG 682
 QY 1321 ctccattgagacagagatgagcaacagatgctgctatgacaacctgatacaacagcc 1380
 DB 683 CTCTATTGAGACACAGAGATGCAACAGCATATGCTATGACAACTGATCACACAGCC 742
 QY 1381 atgactgtgacggcttccctgcaggggaaacgtgattctctgcaggggtgacagtgagg 1440
 DB 743 ATGATGTGGCGGCTTCCGCGAGGGGAACGTGATTTCTTGCAGGGGTGACATGAGAGG 802
 QY 1441 cctctgtcactctcgaagaacataatctgtgtgctgctaggggatacaagctggggtctc 1500

DB 803 CMTCTGCTCACTTCGAAAGCAATCTGCTGCTGATGAGGATACCAAGCTGGGTTCT 862
 QY 1501 gctctgcaaaagcttaccagacagagagtgtaaggaatgtgattatcaagagctg 1560
 DB 863 GCGCTGCCAAAGCTTACACACAGAGATGATGAGGAATGTGATTTACAGGACTG 922
 QY 1561 attatcgacaatgagggcagagcggttaatacatgctctgctctgacgtctt 1620
 DB 923 ATTATGACAAATGAGGGCAGAGCGCTATCCAGATGCTGCTGCTGAGCTGTT 982
 QY 1621 tacaagaacacatgagggctgttctgctcccccggagatattctctagaagatg 1680
 DB 983 TACAGAAAACATGGGCTGCTGCTTCCGCCGATGATTTACTTTAGAGATGAT 1042
 QY 1681 tcagagctcactcatcttattcaacagtgaaact 1715
 DB 1043 TCAGAGTCACTTCTTTTATTTAAACAGTAACT 1077

RESULT 13

US-09-016-434-963
 Sequence 963, Application US/09016434

GENERAL INFORMATION:
 APPLICANT: Janice Au-Young
 TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
 TITLE OF INVENTION: PATHWAY GENE EXPRESSION
 NUMBER OF SEQUENCES: 1490
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
 STREET: 3174 PORTER DRIVE
 CITY: PALO ALTO
 STATE: CALIFORNIA
 COUNTRY: USA

ZIP: 94304
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/016,434
 FILING DATE: HERewith
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:

FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Zeller, Karen J.
 REGISTRATION NUMBER: 37,071
 REFERENCE/DOCKET NUMBER: PA-0002 US

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (650) 855-0555
 TELEFAX: (650) 845-4166
 INFORMATION FOR SEQ ID NO: 963:

SEQUENCE CHARACTERISTICS:
 LENGTH: 1001 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear

IMMEDIATE SOURCE:
 LIBRARY: PROSTUT05
 CLONE: 842889
 US-09-016-434-963

Query Match 50.7%; Score 881.2; DB 34; Length 1001;
 Best Local Similarity 95.2%; Pred. No. 2e-248;
 Matches 936; Conservative 9; Mismatches 31; Indels 7; Gaps 4;

QY 21 gcgagggcgggagcgccgctgga--gcgcgagagtcataatgacattcagaba 77

```
; EARLIER APPLICATION NUMBER: 60/125,469
; EARLIER FILING DATE: 1999-03-22
; EARLIER APPLICATION NUMBER: 60/125,560
; EARLIER FILING DATE: 1999-03-22
; NUMBER OF SEQ ID NOS: 1296
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 1044
; LENGTH: 1181
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1181)
; OTHER INFORMATION: n = A,T,C or G
US-09-526-996-1044
```

```
Query Match      26.7%; Score 463.6; DB 92; Length 1181;
Best Local Similarity 98.9%; Pred. No. 2,2e-125;
Matches 466; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 481 ggtacctgcatcaacccttaactggtgtgatggcgttcaacactgcccgcggggag 540
Db 711 ggtacctgcatcaacccttaactggtgtgatggcgttcaacactgcccgcggggag 770
QY 541 gacgagaatcgtgtgtgttcgctcttaacgacaaacttcaccttcaggttactatct 600
Db 771 gacgagaatcgtgtgtgttcgctcttaacgacaaacttcaccttcaggttactatct 830
QY 601 cagagagaagtcctcgacacctgtgtgccaagaagactggaacgaaactacggggcg 660
Db 831 cagagagaagtcctcgacacctgtgtgccaagaagactggaacgaaactacggggcg 890
QY 661 gacctgacggagacatgggctataagaataattttactcttagccaaggaataatggaatgac 720
Db 891 gacctgacggagacatgggctataagaataattttactcttagccaaggaataatggaatgac 950
QY 721 agcgatccacacagcttatagaactgaaacaaagtgcggcaatgtcgaatatataaa 780
Db 951 agcgatccacacagcttatagaactgaaacaaagtgcggcaatgtcgaatatataaa 1010
QY 781 aaactgtacacacagtgatgctgttcttcaaaagcaatgttcttcaagctgtatagcc 840
Db 1011 aaactgtacacacagtgatgctgttcttcaaaagcaatgttcttcaagctgtatagcc 1070
QY 841 tggcggtgtcaactgtgaactgaagccggcagagagatgttgggcgggcgaagagcgctc 900
Db 1071 tggcggtgtcaactgtgaactgaagccggcagagagatgttgggcgggcgaagagcgctc 1130
QY 901 ccgggggcttgcccttgcccttgcaagtcagcgtccaggtccagaacgtccacgtg 951
Db 1131 ccgggggcttgcccttgcccttgcaagtcagcgtccaggtccagaacgtccacgtg 1181
```

Search completed: April 22, 2000, 06:15:46
Job time: 2285 sec

D	417	ATCCAGTGTGTGTGGCAACCCAGGCAATGATGTGACACAGAAACAGTGTGCTGATT	358
Q	1255	TCGGTGTGGTGGGCGCACACCGAGAGAGAAAGGAGAACCTCGAAGTGTCTGAAGTCTCGAAG	1334
D	357	TCGGGGTGGGGGGCCACACCGAGAGAAAGGAGAACCTCAAAGTGTCTGAACCTTCNCAAG	298
Q	1315	GTGCTCTCATTCTTAAGACACAGAGATGCAACAGTCAGTATGTCTTAAGACAACTGTATCA	1374
D	297	GTGTCTTCTCATTAGACACAGAGATCTCAACAGAGATGTTATGTATGACAACTGTATCA	238
Q	1375	CCAGACATGATCTGTGTGCGGCTCTCCGCGAGAGAGAGTCTGATCTCTCTCGAAGTGTACGT	1434
D	237	CCACCCATGATCTGTGTGCGGCTCTCTGCGAGAGAGAGAGTGTGATCTTTCGACGGGTACAGT	178
Q	1435	GGAGGAGCTCTGTCTCACTCTGAGAGAACCAATATCTGTGTCTGATAGAGAGGATAAGGCTGTG	1494
D	177	GGAGGGCTCTGGGTCCCTTTCGAGAAACAATATCTGTGTCTGATAGGAGATACAAGCTGG	118
Q	1495	GGTCTGTGTCTGTCCAAAGCTTACAGACCCAGAGTGTACGGGAATGTGATGTATCCCG	58
D	117	GGTCTGTGTCTGTCCAAAGCTTACAGACCCAGAGTGTACGGGAATGTGATGTATCCCG	58
Q	1555	GACTGGTTATATCGACAAGATGAGAGAGAGAGATATCGACATGTGCTTCGCTCTTG	1611
D	57	GACTGGATTATATCGACAAGATGAGAGAGAGAGATATTCACATGATGTTTCGCTCTTG 1	

RESULT		4	
A199453/C			
LOCUS	A1994553	385 bp	mRNA
DEFINITION	t080605.x1 NCI CGAP CCL1 Homo sapiens CDNA clone IMAGE:2108144 3		04-FEB-1999
ACCESSION	G141149 CO TR:O15393 O15393 SERINE PROTEASE.		, mRNA sequence.
VERSION	A1994553		
KEYWORDS	A1994553.1 GI:4224100		
SOURCE	Est.		
	human.		

Eukaryota; Metazoa; Chordata; Craniata Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homidae; Homo;
 1 (bases 1 to 385)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
 Tumor Gene Index
 Unpublished (1997)
 On Mar 9, 1998 this sequence version replaced gi:2946770.

Tel.: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever, M.D., Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/MLN at: www.bio.lnlnl.gov/dbtrp/Image/Image.html

```

Trace considered overall poor quality
Seq primer: -40UP from Gibco
High quality sequence stop: 1.
FEATURES
  Location/Qualifiers
    source             1..385

```

[illegible]

adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized, and was constructed by Bento Soares and M. Fatima Bonaldo."

Query Match	21.3%	Score	370.6	DB	45	Length	385
Best Local Similarity	97.7%	Pred. No.	5.2e-91				
Matches	376	Conservative	0	Mismatches	9	Indels	0
						Gaps	0

QY	1201	ggtgtctctgcacaaaccacagga	ctgaatgctgcagccagaaacagctctgcgtatctccagg	1260
Db	385	gggtgtctgcacaaaccacagga	ctaatgtgtacacccaaacagctctgcgtgaatttccggg	326
QY	1261	tggggggccacacgagaggaagaa	gagaaacctcagaagctgcagagctgcgcaagtgtcct	1320
Db	325	tggggggggccacacgagaggaagaa	ggaaagacctcaaaagtctcgaacctccaaagggtgctt	266
QY	1321	ctcaattgagacacagagatgtcaca	agcagatatgtctctatgacaaaccttggttcacacagcc	1380
Db	265	ctcattttaaACACAGAGAAgCAAC	ACAATAgtTtATgACAAcCTGATCACACCAgCC	206
QY	1381	atgattctgtgcacgagctctcctgc	aggggaacgtctgactctctgcacagggtgcacagtggag	1440
Db	205	atgatttTtTgCGGgctTCTCTGc	AGGgAAAGtCGATtTCTTgCCAGGgTGAcAOTGCAgGc	146
QY	1441	cctctgctactctcgaagaacaata	ctctgtgtctatagaggatacaagctcggggtctc	1500
Db	145	cctctgctactCTTcGAAgAAcAA	TATtTtTGTgTgGCTATgAGGgATtCAAAcCTGGGgtTCT	86
QY	1501	gagctgtgcacaaagcttaacaga	ccaggaagctgaaggaaatgtgattcatcaaggagctgg	1560
Db	85	ggctgtgCCAAAGcCTTACAGAG	ACCAGAGAGtGTAAGGTATtTCCACAGAGCTGg	26
QY	1561	atttatcgacaaatgagggcacag	g	1585
Db	25	atttatCGACAAATgAGGGcACAG	c	1

RESULT	5
AIR61741/C	
LOCUS	A1261741 353 bp mRNA
DEFINITION	qz3id04.x1 NC_IGAP_K1d11 Homo sapiens cDNA EST
	similar to TR:O15393 O15393 SERINE PROTEASE. ; , mRNA sequence.
ACCESSION	A1261741
VERSION	A1261741.1 GI:3869944
SOURCE	EST.
ORGANISM	human.
	Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota, Metazoa;
Eutheria; Primates;
Carnivora; Homnidae; Homo.
1 (bases 1 to 353)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On Jan 14, 1998 this sequence version replaced gl:1797896.

Tel: (301) 496-1550
Email: Robert.L.Strausberg@nih.gov
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert Length: 271 Std Error: 0.00
Seq primer: -40UP from Gibco

Query Match 17.7%, Score 307.2, DB 47, Length 564;

[illegible]

```

OY 308 TAFAGILQSFH-FYGAQYQYEVKYSHPNYSKKNNDIAKMLKLPFTFNDLVKPCVLP 366
DB 122 GPGPRLATGTCWVNGLVHSGEALAS-VLQEVAVPLDSDNMCMLYLGSEPLAGRLI 180
OY 367 NPGMMLQPEQLCWSGWCATEKTEKTSVLAAYKVLLETORCNSRYV-D-NL-----I 419
DB 181 QDDMLCAGSVGKDKSCGDSGGLPCPINDTWIAGIVSMGFCARFPGVYQVLSY 240
OY 420 TPAMICAGFLGNDVSCGDSGGLVTSKNNIMWLIGDTSMGSCAKAYRGVGNVWF 479
DB 241 TDMTQRTL 248
OY 480 TDMTQRTM 487

RESULT 13
ID W77302 standard; Protein; 271 AA.
AC W77302;
DE Amino acid sequence of SP001LA, a homologue of HELA2.
KM Serine protease; regulation; cell activity; viability; HELA2; ATC2;
KM BOM3; testis; fertility; suppressor; testicular germ cell cancer;
KM Seminoma; testis-specific expression; antitumour; sperm development;
KM Infertility; human; chromosome 16p13.3.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Disulfide_bond 29 /note- "likely to be involved in disulphide bonding"
FT Misc_difference 44 /note- "likely to be a catalytic residue"
FT Disulfide_bond 45 /note- "likely to be a catalytic residue"
FT Misc_difference 95 /note- "likely to be involved in disulphide bonding"
FT Disulfide_bond 115 /note- "likely to be a catalytic residue"
FT Disulfide_bond 129 /note- "likely to be involved in disulphide bonding"
FT Misc_difference 160 /note- "likely to be involved in disulphide bonding"
FT Disulfide_bond 162 /note- "encoded by GGAA"
FT Disulfide_bond 181 /note- "likely to be involved in disulphide bonding"
FT Disulfide_bond 192 /note- "likely to be involved in disulphide bonding"
FT Disulfide_bond 196 /note- "likely to be involved in disulphide bonding"
FT Misc_difference 202 /note- "likely to be a catalytic residue"
FT Disulfide_bond 220 /note- "likely to be involved in disulphide bonding"
FT Misc_difference 235 /note- "likely to be involved in disulphide bonding"
PN W09836054-A1.
PD 20-AUG-1998.
PR 13-FEB-1998; A00085.
PR 18-NOV-1997; A0-000422.
PR 13-FEB-1997; A0-005101.
PA (AMRA-) AMRAD OPERATIONS PTY LTD.
PI Antalis TM, Hooper JD;
DR WPI: 98-480768/41.
DR N-PSDB; V59134.
PT New serine protease(s) and kinase involved in regulating cell
PT activity and viability - particularly the testis-specific protease
PT HELA2 used for modulation of fertility and as tumour suppressor
PS Example 15; Fig 20A; 167pp; English.
CC W77302-04 represent HELA2 homologues. The genes are found in a cluster
CC on chromosome 16p13.3. HELA2 was isolated from HeLa cells, and has
CC homology to serine proteases. The protein is involved in or associated
CC with regulation of cell activity and/or viability. Administration of
CC recombinant HELA2 (also called testisin) is used to increase fertility.

```

```

CC Downregulation of HELA2 reduces fertility. HELA2 is also a suppressor of
CC testicular germ cell cancers (seminoma) and is also expressed in some
CC non-testicular cancers (of colon, pancreas, prostate and ovary), so is
CC a marker/potential therapeutic target for cancer. The promoter from the
CC HELA2 gene is useful for testis-specific expression of other genes,
CC e.g. for gene therapy or modulation of fertility. Drugs that block
CC activity of HELA2 should have antitumour activity (other than in
CC testis) while in testis recombinant HELA2 should stop growth of tumours
CC and normalise sperm development (eliminating the need for orchidectomy).
CC Identification of mutant forms of HELA2 can be used to diagnose
CC infertility.
SO Sequence 271 AA;

Query Match 16.6%; Score 608; DB 1; Length 271;
Best Local Similarity 37.4%; Pred. No. 5,806-48;
Matches 89; Conservative 54; Mismatches 88; Indels 7; Gaps 5;

DB 2 NRWVGEDSTSEMPWIVSIQKNGTHHCAGSLTSRWVITAHACEKDNLPFLSVLGG 61
OY 253 SRVGGESALPGAMPQVSLHVOVNVHCGSITPEWIVTAHCVEKPLNPMWHTAFAG 312
DB 62 AMQGNPESRSQKQVAVPEHPVYSWKEGACADIALVRLERSIOFSRVLPICLPDASI 121
OY 313 ILROSEMFY-GAGYQVEKVIHPNDSTKKN-NDIALKRLQKLPFTFNDLVKPCVLP 370
DB 122 HLPPTHCMISGWSIQDGVPLPHPOTLQKLVPIIDSEVCSHLYWRGAGGPIEDMJC 181
OY 371 MLOPEQLCWSGWCATEKGR-T-SEVLAAYKVLLETORCNSR- --VYDNLITPAMIC 425
DB 182 AGYLEGERDACLDGSGGLPCLMOCVDAWLLAGIISWEGCARNRPGVYSLSAHSRW 239
OY 426 AGFLGNDVSCGDSGGLVTSKNNIMWLIGDTSMGSCAKAYRGVGNVWFMDI 483

RESULT 14
ID Y13391 standard; Protein; 317 AA.
AC Y13391;
DE Amino acid sequence of protein PRO343.
KM Secreted protein; transmembrane protein; human; enterocolitis;
KM Zollinger-Ellison syndrome; gastrointestinal ulceration;
KM congenital microvillus atrophy; skin disease; cell growth;
KM abnormal keratinocyte differentiation; psoriasis; epithelial cancer;
KM Parkinson's disease; Alzheimer's disease; ALS; neuropathy;
KM fibromodulin; dermal scarring; Usher Syndrome; Atrophla areata;
KM anti-Chromobac; wound healing; tissue repair.
OS Homo sapiens.
PN W09914328-A2.
PD 25-MAR-1999.
PR 16-SEP-1998; U19330.
PR 15-NOV-1997; US-066840.
PR 17-SEP-1997; US-059113.
PR 17-SEP-1997; US-059115.
PR 17-SEP-1997; US-059117.
PR 17-SEP-1997; US-059119.
PR 17-SEP-1997; US-059121.
PR 17-SEP-1997; US-059122.
PR 17-SEP-1997; US-059184.
PR 18-SEP-1997; US-059263.
PR 18-SEP-1997; US-059266.
PR 15-OCT-1997; US-062125.
PR 15-OCT-1997; US-062126.
PR 17-OCT-1997; US-062285.
PR 17-OCT-1997; US-062287.
PR 21-OCT-1997; US-063486.
PR 21-OCT-1997; US-062814.
PR 24-OCT-1997; US-062816.
PR 24-OCT-1997; US-063045.
PR 24-OCT-1997; US-063120.
PR 24-OCT-1997; US-063121.
PR 24-OCT-1997; US-063122.
PR 24-OCT-1997; US-063127.
PR 24-OCT-1997; US-063128.
PR 27-OCT-1997; US-063329.
PR 27-OCT-1997; US-063327.

```

[illegible]

QY	332	ggaacagctgctaccctcaagaagcaagaagcaactgtgcatcacccttcgaaccccttgggagac	391
Dd	69	gagcactgctacccttctaaagtctaaagaaatccctgtctgttttagccctccctgggacctg	128
QY	392	tccctgtggagctgtgcctgcgcgcctgtgcctactctctggaaagttcattggtgcaagaagtgc	451
Dd	129	tcctcacaggagctgctgtgagctgctgctgtctgtctgtttggaggttttgaggacagcaactg	188
QY	452	ccaactctggatagaagctgcgaactctcaagtgatcgtatcgaacccctctaacctggtgtg	511
Dd	189	ctacgcttgtagaagagagtggtggctcttcacggacacatgcacgtccttctctggtgtg	248
QY	512	atggcgtgctgacacgtcccccggcgggggaagaaagatcgtgtgttctgacctcaagac	571
Dd	249	acggggtagacacttctccacacggaagaaagagacacacgttctgttctcttccagac	308
QY	572	caaaactcatccttcagagtgatctactctcacaaggaagtcctgtgacccctgtgtgcaag	631
Dd	309	aaagcttcatctccacaggtttactatcttcagaggaagacccgtgacccgtggtccag	368
QY	632	acgaactggaacgaagaaactacagggcgggcgctgcgaaggacatggtgcataagaataa	689
Dd	369	atgattgtagatgagactctacggagagacgacatgtaagacatgaggaatacgaacaa	426
RESULT	11		
LOCUS	AA244269/c		
DEFINITION	AA244269	259 bp	EST
ACCESSION	AA244269		20-AUG-1997
VERSION	AA244269.1		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS	Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
TITLE	1 (bases 1 to 259)		
JOURNAL	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .		
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
	Unpublished (1997)		
	Contact: Robert Strausberg, Ph.D.		
	Tel: (301) 496-1550		
	Email: Robert.Strausberg@nih.gov		
	Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuagui, M.D., Michael Emmert-Buck, M.D., Ph.D.		
	cDNA Library Preparation: David B. Kitzman, Ph.D.		
	cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.		
	DNA sequencing by: Washington University Genome Sequencing Center		
	clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/URL at:		
	www.bio.lnlni.gov/dbfp/image/image.html		
FEATURES			
SOURCE			
	1.259		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:1007327"		
	/clone_1lb="NCI-CGAP_Pr1"		
	/sex="Male"		
	/dev_stage="45 years old"		
	/lab_host="DH10B"		
	/note="Vector: PAMP10; site_1: NotI; site_2: EcoRI; 1st strand cDNA was primed with oligo(dT)17 on 50 ng of DNase-treated, total cellular RNA obtained from 5,000-10,000 microdissected, histologically normal, prostate epithelial cells. Double-stranded cDNA was ligated to EcoRI adaptors, 5 cycles of PCR applied to the		

Search completed: Wed Apr 19 23:28:04 2000
Job time : 312 secs.

```

OY 113 CSNSGIECCSSGTCINPSMWCDSVSHCPGEGEDNRCVRLY-GP-NFI-LQVYSSQKSMH 169
DB 480 VACAEWMTQISDDVQCLGLGTG-NSVPTFS-TGGGPRVINTAP-NSLLTL-PSQO 535
OY 170 PVCQDDMNENYGAARCRDYGKNNFYSSQIYDDSGTSTPMKINTSAGNDVTKKTHSDA 229
DB 536 CLEDSILLQCNKSCGKKLVTOEVPKIVGSDSREGAPMVVAALYFDDQVCGASLVS 595
OY 230 CSKAVVSLRCT--ACGVNLNNSROS-RIVGGSALPGAMPQVSLHVQNVHVCSSIT 286
DB 596 RDLVSAHCYVGRNME-PSKMAVLGLHMASLSPQIETRLIDIVINPHYNKRKKN 654
OY 287 PEWIVAAHCV-EKPLNPMHMTAFAGILRQSMFYGA-GYQ-VERKVISHPNYSKTKNN 343
DB 655 DIAMHLEKVNVTIDYIOPICLPEENOVFPGRICISAGCALIYOGSTADVQLEADVPL 714
OY 344 DIAMKLRPLTENDLVKPCVLPNPGMILQPEQLCWSIGWATEEKGKSEVLNAKAVLL 403
DB 715 LSNKCCQOQ-MPEYNITENNVACAGYAGVSCQSGDGPGLMCQENRMILAGVTSFGYQ 773
OY 404 IETORNSRKYVNLITPMICAGFLQGNVDSQSGDGPGLVSKNNIMWLIIDTISWGS 463
DB 774 CALPNRPYARVPRFTWI 793
OY 464 CAKAYRPGYGVNMFVTDWI 483

```

```

RESULT 2
ID W22986 standard; Protein: 233 AA.
AC W22986;
DT 08-OCT-1997 (first entry)
DE Human serine protease 60 (SP60).
KM Human; colon carcinoma; COLO 201; cell line; serine protease; SP60;
  screening; inhibitor; treatment; disease.
PN Homo sapiens.
OS 10-JUN-1997.
PD 24-JUL-1996; 212196.
PR 29-SEP-1995; JP-275105.
PA (SUNR) SUNTORI LTD.
DR MPI; 97-357902/33.
N-PSDB; T79127.
PT Human colon carcinoma derived serine protease(s) SP59, SP60 and SP67
  study agent for treatment of various diseases
PS Claim 1; Pages 11-12; 16pp; Japanese.
CC The present sequence is the human colon carcinoma COLO 201
  cell line derived serine protease 60 (SP60), which can be used to
  screen for specific inhibitors, e.g. to search for, or study an
  agent for the treatment of various diseases.
SQ Sequence 233 AA;

```

```

Query Match 19.3%; Score 710; DB 1; Length 233;
Best local Similarity 44.6%; Pred. No. 2,36e-58;
Matches 104; Conservative 48; Mismatches 71; Indels 10; Gaps 9;

```

```

DB 1 VVGGEASVDSMPQVSIQYDKOHVCGSILDRHVVITAAHCRKH-TDVPNKKVRAQSD 59
OY 255 IYGGESALPGAMPQVSLHVQNVHVCSSITPEWITAAHCVKPLNNPMHMTAFAGIL 314
DB 60 KLG-SF--PSLAIVAKII-TEFNPMYKNDIALMKLOPLTFSGVTRPCLPFDEDEL 115
OY 315 ROSMFAGAGYQVEKVISHPNYSK-TKNDIALMKLOKPLTFDVLKPCVLPNPGMIL 373
DB 116 PATPL-WIIGMFTKONGKMSDILLOASVOVIDSTRCNADAVGTEKMKAGIPEG 174
OY 374 PE-QLCIVSGMGAEEKG-KTSEVLNAKAVLLIETORNSRYVDNLTIPAMICAGFLQ 431
DB 175 GVDTGSDSGPLMYQSDQ-WHVVGIVSKVGGCGPSTPGYTVVSAVLNMIY 226
OY 432 NVDSQDSDSGPLVTSKNNIMWLIIDTISWGSCKAKAYRPGYGVNMFVTDWI 484

```

```

RESULT 3
ID R89435 standard; Protein: 418 AA.
AC R89435;
DT 27-MAY-1996 (first entry)
DE Trypsin-like enzyme.
KM Trypsin; fibrinogen; thrombin; expectorant; respiratory disease;
  asthma; VIP; vasoactive intestinal peptide; influenza virus;
  protease; primer; PCR; amplification.
OS Homo sapiens.
FH Key
FT Peptide 1..186
FT Protein /label_sig_peptide
FT /label_mat_protein
PN AU9527248-A.
PD 08-FEB-1996.
PR 31-JUL-1995; 027248.
PR 29-JUL-1994; JP-178607.
PA (TEIJ) TEIJIN LTD.
PI Masuda K, Ogawa H, Suga T, Sugimoto Y, Takagi K;
  Yamoka K, Yasuoka S;
  MPI; 96-117356/13.
DR N-PSDB; T10689.
PT Nucleic acid sequence encoding trypsin-like enzyme - which digests
  fibrinogen, used as expectorant in treatment of respiratory
  diseases, e.g. bronchial asthma
PS Example 11; Page 47-49; 65pp; English.
CC The overlap parts of 107 bp between the sequences given in
  T10698 and T10703 were identified, and thereby their identity
  was confirmed. It was confirmed from the sequence analysis,
  CC that these overlapping sequences contained a region encoding the
  CC amino acids of the N-terminus 20 residues of the trypsin-like
  CC enzyme isolated from the cough phlegm. The sequences were
  CC ligated, and the desired trypsin-like enzyme gene cDNA sequence
  was determined (T10689).
SQ Sequence 418 AA;

```

```

Query Match 19.0%; Score 697; DB 1; Length 418;
Best local Similarity 42.3%; Pred. No. 5.04e-57;
Matches 105; Conservative 49; Mismatches 87; Indels 7; Gaps 6;

```

```

DB 173 CGA-GPDLITLSEORILGTEAEESMPQVSLRLNNAHCGSLINNMILTAHCF-R 230
OY 240 CIACGVNLNNSRQSRIVGGSALPGAMPQVSLHVQNVHVCSSITPEWITAAHCVK 299
DB 231 SNSNPRDWTSGI-STTP--KLRRKRVNILLNNYKSKTHENDALVRLNSVFTFD 287
OY 300 PLNNPMHMTAFAGILRQSMFYAGYQVEKVISHPNYSKTKNDIALMKLOKPLTFNDL 359
DB 288 IHSVCLEPAATQNIIPGSTAVVTGMAOEYAGHTVPELRQGVRIISNDVCNAPHSYNGAI 347
OY 360 VKFVCLPNEGMILOPEQLCHISGMGAEEKTSEVLNAKAVLLIETORNSRYVDNLI 419
DB 348 LSGMTCAGVPGGVDAQCGSDGGLVQEDSRRLMFTVGVISMGDGLPDKPGVYTRVTA 407
OY 420 TPAMICAGFLQGNVDSQSGDGPGLVTSKNN-IMWLIIDTISWGSCKAKAYRPGYGVN 478
DB 408 YLDWI-RQ 414
OY 479 FTDWIYRQ 486

```

```

RESULT 4
ID W46917 standard; Peptide: 356 AA.
AC W46917;
DT 02-JUL-1998 (first entry)
DE Amino acid sequence of a novel human kallikrein.
KM Kallikrein; HKLP; human; serine protease; drug screening; atagonist;
  agonist; treatment; hypertension; cardiac hypertrophy; arthritis;
  inflammatory disorder; blood clotting disorder.
OS Homo sapiens.
FH Key
  Location/Qualifiers

```

```

DEFINITION      t909d11.x1 NCI-CGAP-CLL1 Homo sapiens cDNA clone IMAGE:2108277 3'
ACCESSION       AI393270
VERSION          AI393270.1 GI:4222817
KEYWORDS        EST.
SOURCE           human.
ORGANISM         Homo sapiens
                  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
                  Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE        1 (bases 1 to 450)
                  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS          National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
                  Tumor Gene Index
TITLE            Unpublished (1997)
JOURNAL          On Jan 19, 1998 this sequence version replaced gi:2153106.
COMMENT          Contact: Robert Strausberg, Ph.D.
                  Tel.: (301) 496-1550
                  Email: Robert_Strausberg@nih.gov
                  Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
                  M.D., Louis M. Staudt, M.D., Ph.D.
                  CDNA Library Preparation: M. Bento Soares, Ph.D.
                  CDNA Library Arrayed by: Greg Lennon, Ph.D.
                  DNA Sequencing by: Washington University Genome Sequencing Center
                  Clone distribution: NCI-CGAP clone distribution information can be
                  found through the I.M.A.G.E. Consortium/LLNL at:
                  www.bio.llnl.gov/bdrrp/Image/Image.html

FEATURES
   source
       1..450
           /organism="Homo sapiens"
           /db_xref="taxon:9606"
           /clone="IMAGE:2108277"
           /clone_1id="NCI-CGAP-CLL1"
           /tissue_type="B-cell, chronic lymphocytic leukemia"
           /lab_host="DH10B"
           /note="Vector: pT73D-Pac (Pharmacia) with a modified
           polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
           was primed with a Not I - oligo(dT) primer [5',
           TGTTCACCAATCTGAAGTGGGCGCGCGATGCTTTTATTTTATTTTATTTT
           T 3']; double-stranded cDNA was ligated to Eco RI
           adaptors (Pharmacia), digested with Not I and cloned into
           the Not I and Eco RI sites of the modified pT73 vector.
           Library is normalized, and was constructed by Bento
           Soares and M. Fatima Bonaldo."

BASE COUNT      98 a      132 c      107 g      111 t      2 others
ORIGIN
Query Match      25.7%; Score 446.4; DB 45; Length 450;
Best Local Similarity 99.3%; Pred. No. 9.1e-112;
Matches 447; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1162 ctgcgaagacctgacttcaagacctagtgaacacagtgctgtcccaaccaggc 1221
      |||||||
DB 450 CTGCAGAGCCTCTGACTTTTCANCGACTAGTGAACACAGTGTCTGCCCAACCAGGC 391

QY 1222 atgattctgcagcagcagcagctctgcgtgattctcgggttgggggccacagagagaa 1281
      |||||||
DB 390 ATGATGCTGCAGCCGACGACAGCTCTGCTGNNATTTCCGGGTGGGGGCCACGAGAGAAA 331

QY 1282 gggaagacctcagaagtgctgaacgctgcgaagtgcttcattcagacacagagatgc 1341
      |||||||
DB 330 GGGAAGACCTCAGAAAGTCTGTAACGCTGCCAAAGTGCTTCATATGAGACACAGAGATGC 271

QY 1342 aacgcagatatgtctatgacaacctgtatcacaccagcagatctgtggcgctctctg 1401
      |||||||
DB 270 AACGACAGATATGCTATGACAACTGATCACACACGACGATGATGTGGCGGCTTCCTG 211

QY 1402 cagggaacgctcagatctctgcagggtcacagtgtaggggcccctgtcaactcgaagaac 1461
      |||||||

```

```

DB 210 CAGGGAACGTGATCTTGTCCAGGCTGACAGTGGAGGCCCTCTGCTACTTCAGAGAAC 151
QY 1462 aatctctgtgcgtatagaggatatacaagctcgaggctctgtgcgaagcttacaga 1521
      |||||||
DB 150 AATATCTGTGCTGATATGAGGAGATCAAGCTGGGGTTTGCTGTGCTGCCAAAGCTTACGA 91

QY 1522 ccaggagtgctacggagatgtatgtattcaccgagctgagattatcgaacaatgaggca 1581
      |||||||
DB 90 CCAGAGAGTACGGGAATGATGATATTACAGCAGCTGATTTATCGCAATGAGGCCA 31

QY 1582 gacggtatccacacatgtcttcgctctg 1611
      |||||||
DB 30 GACGCTAATCCACATGGTCTTCCTCTTG 1

RESULT 3
LOCUS      AI673506 417 bp. mRNA EST 18-MAY-1999
DEFINITION w19h04.x1 Soares_Dieckgraefe_colon_NHUC Homo sapiens cDNA clone
IMAGE:2351095 3' similar to SW:TMS2_HUMAN O15393 TRANSMEMBRANE
PROTEASE, SERINE 2 ;, mRNA sequence.
ACCESSION  AI673506
VERSION     AI673506.1 GI:4853237
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE    1 (bases 1 to 417)
              NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
              Tumor Gene Index
TITLE        Unpublished (1997)
JOURNAL      On May 18, 1998 this sequence version replaced gi:3138221.
COMMENT      Contact: Robert Strausberg, Ph.D.
              Tel.: (301) 496-1550
              Email: Robert_Strausberg@nih.gov
              This clone is available royalty-free through LLNL; contact the
              IMAGE Consortium (info@image.llnl.gov) for further information.
              Seq primer: -400P from glbco
              High quality sequence stop: 215.
              Location/Qualifiers
                  1..417
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:2351095"
                      /clone_1id="Soares_Dieckgraefe_colon_NHUC"
                      /tissue_type="colonic mucosa from 5 ulcerative colitis
                      patients"
                      /lab_host="DH10B (phage-resistant)"
                      /note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
                      modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st
                      strand cDNA was primed with a Not I - oligo(dT) primer [5',
                      TGTTCACCAATCTGAAGTGGGCGCGCGATGCTTTTATTTTATTTT
                      T 3']; double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Not I and cloned into the Not I
                      and Eco RI sites of the modified pT73 vector. Library
                      went through one round of normalization. Tissue samples
                      provided by Dr. Brian Dieckgraefe (Washington University,
                      dieck@im.wustl.edu); colonic mucosa represents a range of
                      disease involvement from mild cryptitis to severe
                      ulceration, fibrosis, and degeneration. Library
                      constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT      90 a      125 c      98 g      103 t      1 others
ORIGIN
Query Match      22.7%; Score 395.2; DB 50; Length 417;
Best Local Similarity 96.6%; Pred. No. 9.6e-98;
Matches 403; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1135 aaacagtgctctgcacacccagcagcatgatgtcagcagagacagctctgctgatt 1234
      |||||||

```


PR 28-OCT-1997; US-063541.
 PR 28-OCT-1997; US-063542.
 PR 28-OCT-1997; US-063544.
 PR 28-OCT-1997; US-063549.
 PR 28-OCT-1997; US-063550.
 PR 28-OCT-1997; US-063564.
 PR 29-OCT-1997; US-063435.
 PR 29-OCT-1997; US-063704.
 PR 29-OCT-1997; US-063732.
 PR 29-OCT-1997; US-063738.
 PR 29-OCT-1997; US-063734.
 PR 29-OCT-1997; US-064215.
 PR 29-OCT-1997; US-063735.
 PR 31-OCT-1997; US-063870.
 PR 31-OCT-1997; US-064103.
 PR 03-NOV-1997; US-064248.
 PR 07-NOV-1997; US-064809.
 PR 12-NOV-1997; US-065186.
 PR 17-NOV-1997; US-065846.
 PR 18-NOV-1997; US-065693.
 PR 21-NOV-1997; US-066120.
 PR 21-NOV-1997; US-066354.
 PR 24-NOV-1997; US-066772.
 PR 24-NOV-1997; US-066466.
 PR 24-NOV-1997; US-066770.
 PR 24-NOV-1997; US-066511.
 PR 24-NOV-1997; US-066453.
 PR (GEM) GEMENTECH INC.
 PI Chen J, Goddard A, Gurney AL, Pennica D, Wood WI, Yuan J;
 PI WPI: 99-229533/19.
 DR N-PSDB; X52262.
 PT New isolated human genes and polypeptides used in, e.g. treatment of
 PT gastrointestinal ulceration
 PS Claim 12, Fig 98, 320pp; English.
 Y1344-403 represent acute and transmembrane human proteins.
 CC The CDNA sequences are obtained from cDNA libraries, prepared from
 CC fetal lung, fetal kidney, fetal brain, fetal liver and fetal retina.
 CC The encoded polypeptides have specific uses based on their homology to
 CC known polypeptides, e.g. PRO211 and PRO217 can be used for disorders
 CC associated with the preservation and maintenance of gastrointestinal
 CC mucosa and the repair of acute and chronic mucosal lesions
 CC (e.g. enterocolitis, Zollinger-Ellison syndrome, gastrointestinal
 CC ulceration and congenital microvillus atrophy), skin diseases associated
 CC with abnormal keratinocyte differentiation (e.g. psoriasis, epithelial
 CC cancers such as lung squamous cell carcinoma of the vulva and gliomas),
 CC potent effects on cell growth and development, diseases related to growth
 CC or survival of nerve cells including Parkinson's disease, Alzheimer's
 CC disease, ALS, neuropathies or cancer. PRO265 can be used as for
 CC fibromodulin, e.g. for reducing dermal scarring. PRO264 can be used
 CC as a target for anti-tumor drugs. PRO533 may be used in the treatment
 CC of Usher Syndrome or Atrophia areata. PRO269 can be used as an
 CC anti-thrombotic agent. PRO287 polypeptides and portions may have
 CC therapeutic applications in wound healing and tissue repair. PRO317 can
 CC be used for treating problems of the kidney, uterus, endometrium, blood
 CC vessels, or related tissue, e.g. in the heart of genital tract.
 SO Sequence 317 AA;

Query Match 16.68; Score 608; DB 1; Length 317;
 Best Local Similarity 37.48; Pred. No. 5,80e-48;
 Matches 89; Conservative 54; Mismatches 88; Indels 7; Gaps 5;

DB 48 NRYVGGDSRDSSEMPVVSLOKNGTHHCAGSLTSRMTVAACFKNLKNKPYLSVLG 107
 Oy 253 SRIYGGESALPGAMPVOSLHVONVHVGSSITPEVITVAACVVERPLNPNMTAFAG 312
 DB 108 AMOLGNFSGSRQKGVAVMEHPVYSWKEGACADIALVRLERSIQSESVLPICLPDASI 167
 Oy 313 ILRQSEMFY-GAGYQVEKVISHPYDSKTKN-NDIALMLKLOKPLTFNDLVKPYCLPBM 370
 DB 168 HLPNTHCMWISGMSIDQVLPHPOTLQKLPYIIDSEVCSHLIYNAGAGQPTDMLC 227
 Oy 371 MLOPEQLCWISGMSGATEKGR-T-SEVLNNAKAYLLIETORNSRY---VYDNLITPMIC 425

DB 228 AGYLEGEDRACIGDSGGPLACVODGMLLAGIISWEGCAENRPGVYISLSAHSRY 285
 Oy 426 AGFLQGNVDSGDSGGGLVYTSKNINIMWLGIDTSWGSAGCAKAYRPGVGNVFTDWT 483

RESULT 15
 ID W64239 standard; Protein: 270 AA.
 AC W64239;
 DT 24-NOV-1998 (first entry)
 DE gerbil homologue of mouse mmp-7 zymogen.
 KW MCP-7 zymogen; gerbil; mast cell protease 7; trypsin-7;
 KW blood clot; anticoagulant; myocardial infarction; reocclusion;
 KW thromboembolism; cerebral embolism; thrombolysis; therapy.
 OS Meriones sp.
 FH Key
 FT Location/Qualifiers
 FT 25..270
 FT /Label= Mat_protein

PN W09824886-A1.
 PD 11-JUN-1998.
 PE 25-NOV-1997; U21620.
 PR 04-DEC-1996; US-032354.
 PA (BGHM) BRIGHAM 6 WOMENS HOSPITAL.
 PI Stevens RJ.
 PI WPI: 98-333308/29.
 DR N-PSDB; V44326.

PT New compositions containing trypsin-7, e.g. mouse mast cell
 PT protease-7 - are used to treat clot formation in e.g. myocardial
 PT infarction, reocclusion following angioplasty or pulmonary
 PT thromboembolism
 PS Disclosure: Page 61, 92pp; English.
 CC This is the amino acid sequence of the gerbil homologue of mouse
 CC mmp-7 (see also W64239), deduced from a
 CC nucleic acid sequence (see V44326). Mature gerbil homologue of
 CC mmp-7 can be used to prevent or inhibit fibrin clot formation in
 CC vitro or in vivo. The invention provides: compositions comprising
 CC an isolated trypsin-7 such as mmp-7, its homologues or humanised
 CC mmp-7; a method for treating a blood clot by administering a
 CC nucleic acid molecule that codes for a trypsin-7, or an expression
 CC product, to decrease fibrinogen activity; a nucleic acid encoding a
 CC serine protease (SP); and a method of producing a mature SP by
 CC expressing the inactive zymogen in a host cell, and cleaving the
 CC enterokinase susceptibility domain. The trypsin-7 polypeptides
 CC can be used to treat disorders mediated by undesirable thrombus
 CC clot formation such as myocardial infarction and reocclusion
 CC following angioplasty of blood clots associated with pulmonary
 CC thromboembolism, deep vein thrombosis, cerebral embolism, renal
 CC vein and peripheral arterial thrombosis. They are also useful for
 CC all surgical procedures that require decreased blood clots.
 SO Sequence 270 AA;

Query Match 16.58; Score 607; DB 1; Length 270;
 Best Local Similarity 40.48; Pred. No. 7.33e-48;
 Matches 99; Conservative 46; Mismatches 80; Indels 20; Gaps 15;

DB 26 IYVGOEA-PGNKWPVOSLPRANETVRRHFGGSLHPQWVLTAAACVCPPIADP-N-KVR 82
 Oy 255 IYVGSALPGA-WPQVSLHVQNV---HVGCSIIITPEVITVAACVVERPLNPNMTAF 310
 DB 83 VQ-LRQVLYIHHLAVSRILTHPTFAONQADTALLEKPNVSISSVHVPSPAS 141
 Oy 311 AGILROSEMFY-GAGYQVEKVISHPYDSKTKNNDIALMLKLOKPLTFNDLVKPYCLPBM 369
 DB 142 EFPFSGFLCVTGMGNINDVSLPPEPLKEVQVNVENOLCDLKYHKGYTGDNIHYR 201
 Oy 370 MLOPEQLCWISGMSGATE-EKGTSEV-LNAARVLLIETORNSRY---VYDNLITPMIC 421
 DB 202 DMLCAQN-BGH-DSCGDSGGPLACVODGMLLAGIISWEGCAENRPGVYISLSAHSRY 259
 Oy 422 A-MICAGFLQGNVDSGDSGGGLVYTSKNINIMWLGIDTSWGSAGCAKAYRPGVGNVFTDWT 480
 DB 260 DWIHR 264
 Oy 481 DWIYR 485

CDNA with an adaptor-specific primer, and the resulting PCR product subcloned into pAMP10 by the UDG-cloning method (Life Technologies). Average insert size is 600 bp. NOTE: Not directionally cloned. This library was constructed by David Kitzman.

BASE COUNT 63 a 77 c 59 g 60 t
ORIGIN

Query Match 13.1%; Score 228.4; DB 30; Length 259;
Best Local Similarity 97.7%; Pred. No. 4.3e-52;
Matches 254; Conservative 0; Mismatches 1; Indels 5; Gaps 2;

Oy 1352 atgtctatgacaacccgtgtccacacagccatgctgtccggcttccttcagaggaacg 1411
Db 255 ACGCTATGACACACTGATTCACACACCCATGATGTGCGGCTTCTGCGGGGAACG 196
Oy 1412 tgcattcttcgacaggtgtacaggtgagggcctctgtcacttcgaagaacaatctgt 1471
Db 195 TCGATTCTTGCCAGGGGTACAGATGAGGGCCCTGTGTCACCTTCGAAGAACAATATCTGGT 136
Oy 1472 ggcctgataggagatacaagctgggggtctctgtctgtccaaagcttaagaccaggaagct 1531
Db 135 GGCCTATGGCGGATCAAGCTGGGGCTCT-GCTGCGCAAACTTACAGACGAGAGTGT 77
Oy 1532 acgggaatgatatgtatccacgagctgattatcgaacaatgagggcagagcgaat 1591
Db 76 ACGGGAATGATGATGATTCACGACGACTGATTTATCGAATAATGAG----GACGCTAAT 21
Oy 1592 ccacatgctctcgtcctgt 1611
Db 20 CCACATGGCTCTGCTCTTG 1

RESULT 12
AA832852 456 bp mRNA EST 23-FEB-1998
LOCUS ub58901.x1 Soares mouse mammary gland NMLMG Mus musculus cDNA clone
DEFINITION IMAGE:1381968 5', mRNA sequence.

ACCESSION AA832852
VERSION AA832852.1 GI:2906580
KEYWORDS EST.

SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 456)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The Mashu-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Jan 9, 1998 this sequence version replaced gi:937909.
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:904436
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 452.

FEATURES
SOURCE

1..456
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:1381968"
/clone_1lb="Soares mouse mammary gland NMLMG"

/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="DB10B"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified polylinker. 1st strand cDNA was prepared from mammary gland tissue from a lactating female, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 99 a 138 c 114 g 105 t
ORIGIN

Query Match 12.9%; Score 223.4; DB 39; Length 456;
Best Local Similarity 71.4%; Pred. No. 1.3e-50;
Matches 309; Conservative 0; Mismatches 121; Indels 3; Gaps 1;

Oy 92 atgctgtgataacagcaagatgtgttgaactcaggtgcacacagcatattgacct 151
Db 27 ATGCTGCTGACACAGCAGATGAGATGCACTGAGGCTACCTCAGATCGACTT 86
Oy 152 actatgaaaaccatgatalaccacccgaaacccatcccgacagccactgtgtcc 211
Db 87 GCTATGAGAACACCGGATATCAGTGTGACACATCTGCTCCGACACAGAGCTC 146
Oy 212 ccactgtctaagagtgatcccggtcagttactaccggtcccggtgccagatgcc 271
Db 147 CCAATGGCTACAACTGTATCAGCCAGTACTACCTTCCTCACTGCTCATGTGCTC 206
Oy 272 cgagggtccctgacgagcgtccaaacccgctgcttcgacgacgacaaatcccatcg 331
Db 207 CGAGGATTCACACGACGACCTTCAATCTGTATCCACACATCCCAAGTCC--TCAG 263
Oy 332 ggcacagttgcaacctcaagaagactaagaagcactgtgatacacttgaaccttggagacct 391
Db 264 GAGCAGCTGTGACACCTCAAGTCAAGATGCTGTGTTAGCCCTCCCGGGGACTG 323
Oy 392 tccctgtggagctgtgctgtgcgtgctactacttggaagtcatgtggcagcaagctct 451
Db 324 TCCCTACGGGAGCTCTGTGCTGCTGCTTGTGGAGGTGTGGAGCAACCTCTT 383
Oy 452 ccaactctggagtagagtgacactcctcaggttactcgtatcaacccctctactgtgtg 511
Db 364 CTACCCCTGAGTGGAGTGGTGTCTTCAGGCATGATGACTCTTCTCTGTGGGTG 443
Oy 512 atggcgtgtcaca 524
Db 444 ACGGGGTAGCACA 456

RESULT 13
A1607826 563 bp mRNA EST 21-APR-1999
LOCUS ub58901.x1 Soares mouse mammary gland NMLMG Mus musculus cDNA clone
DEFINITION IMAGE:1381968 3' similar to TR:015393 015393 SERINE PROTEINASE. ;
IMAGE:1381968 3' similar to TR:015393 015393 SERINE PROTEINASE. ;
mRNA sequence.

ACCESSION A1607826
VERSION A1607826
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 563)
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Ritter,E., Kohn,S., Shin,T., Jackson,T., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.

TITLE The Mashu-HMI Mouse EST Project 1999
JOURNAL Unpublished (1999)


```

Db 326 TGCAGAACCCAGCATGATGCTAGACCTAGACCAAGAAATCTGATTCGGGGGGG 385
Oy 1268 ccaccgagggagaaaggaaagacctcgaagctgtgaaagctgcgaagctcttcac 1325
Db 386 CCACCATATGAGAAAGGAGACCTCGACCTGTGAAATGCTGCCATAGTACCTTGAT 443

RESULT 15
AI504821/c 439 bp mRNA EST 11-MAR-1999
LOCUS AI504821/c
DEFINITION v16/c03.x1 Knowles Solter mouse blastocyst B1 Mus musculus cDNA
; clone IMAGE:976996 3 similar to TR:015393 O15393 SERINE PROTEASE.
; mRNA sequence.
ACCESSION AI504821
VERSION AI504821.1 GI:4402672
KEYWORDS EST
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 439)
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
Ritter,E., Kohn,S., Shih,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
On Apr 7, 1998 this sequence version replaced g1:3034424.
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:557724
This clone was previously sequenced on the 5' end only, this new
data is from the 3' end
Possible reversed clone: similarity on wrong strand
Seq primer: primer name ambiguous
High quality sequence stop: 348.
FEATURES
Source
1. 439
/organism="Mus musculus"
/strain="B6D2 F1/J"
/db_xref="taxon:10090"
/clone="IMAGE:976996"
/clone_1lb="Knowles Solter mouse blastocyst B1"
/tissue_type="blastocyst"
/dev_stage="embryo (pre-implantation)"
/lab_host="DH10B"
/note="Organ: embryo; Vector: pSPORT; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally from mRNA prepared
from 800 blastocysts. Primer: SalI(dT):
5'-CGCTCGACCGCGACCGCTTTTCTTTT-3'. cDNAs were
cloned into the NotI/SalI sites of a pSPORT vector (Life
Technologies). Two different size selections: B1 (larger
inserts) and B3."
BASE COUNT 113 a 111 c 101 g 111 t 3 others
ORIGIN
Query Match 12.2%; Score 212; DB 47; Length 439;
Best Local Similarity 74.0%; Pred. No. 1.7e-47;
Matches 313; Conservative 0; Mismatches 93; Indels 17; Gaps 3;
Oy 1306 gctgcacaggggtctctatgagacacagagatgacagcagatgctctatgacac 1365
Db 439 GCTGCATGTGATCCATCATGATGACGCCCTCAAAATGTATATGATATACATATACANCAAC 380

```

```

Oy 1366 ctgatacacacacacatgatactgtgccgctctctcgaagggaacgtctctccag 1425
Db 379 CTATACACACNAG-CATGATCTGTGCCGCTCTCCAGGGGTCTGTGCACCTTTGCCAG 321
Oy 1426 ggtgacacgttgaagggtcctcgttcacacttcgaagacaaatctcgtgtgacataaggat 1485
Db 320 GGAGACAGTGGAGGGCGCGCTGTTACTTTGAAAGATGGGATCTGTGGCTGATTTGGGGAC 261
Oy 1486 acaagctgggtctgtgtgtgtgccaagcttacagaccagagtgtaagggaatgtgag 1545
Db 260 ACAGACTGGGGCTCGGGCTGTGCCAAGGCACTGAGACTGATATACGGGAACGTGACG 201
Oy 1546 gtattacggaactgattatcgaacaatgaggcgaagcgaatccacatgg----- 1600
Db 200 GTATTACAGATTTGATCTACACGCAAAATGAGGCCAACAACGCTATCCAGCTGGCTTGT 141
Oy 1600 -----ctctgctctgtgacgtgttttacagaataaacaatgggactg-gtttgct 1649
Db 140 CCCAGACTTCTCTTGTCTTCACACACCTTCTGCAGAAACCAAGGGCCCTGAATTTTAC 81
Oy 1650 tcccgctgcatgttacctcttagagatgattcagaaggtcactctcttattataacg 1709
Db 80 TTCTGTGCAACATGTACCTTTTGAGATGTGGAAGGGCGCTTTCACCTTTATTATAACAG 21
Oy 1710 tga 1712
Db 20 TGA 18

```

Search completed: April 22, 2000, 05:45:44
Job time: 2823 sec

1:	/cgn2_6/pdatata/2/pna/PCrnU9_COMB.seq.*
2:	/cgn2_6/pdatata/2/pna/USO80_COMB.seq.*
3:	/cgn2_6/pdatata/2/pna/USO81A_COMB.seq.*
4:	/cgn2_6/pdatata/2/pna/USO81B_COMB.seq.*
5:	/cgn2_6/pdatata/2/pna/USO81C_COMB.seq.*
6:	/cgn2_6/pdatata/2/pna/USO81D_COMB.seq.*
7:	/cgn2_6/pdatata/2/pna/USO81E_COMB.seq.*
8:	/cgn2_6/pdatata/2/pna/USO82A_COMB.seq.*
9:	/cgn2_6/pdatata/2/pna/USO82B_COMB.seq.*
10:	/cgn2_6/pdatata/2/pna/USO82C_COMB.seq.*
11:	/cgn2_6/pdatata/2/pna/USO83A_COMB.seq.*
12:	/cgn2_6/pdatata/2/pna/USO83B_COMB.seq.*
13:	/cgn2_6/pdatata/2/pna/USO84A_COMB.seq.*
14:	/cgn2_6/pdatata/2/pna/USO84B_COMB.seq.*
15:	/cgn2_6/pdatata/2/pna/USO85_COMB.seq.*
16:	/cgn2_6/pdatata/2/pna/USO86_COMB.seq.*
17:	/cgn2_6/pdatata/2/pna/USO87A_COMB.seq.*
18:	/cgn2_6/pdatata/2/pna/USO87B_COMB.seq.*
19:	/cgn2_6/pdatata/2/pna/USO87C_COMB.seq.*
20:	/cgn2_6/pdatata/2/pna/USO88A_COMB.seq.*
21:	/cgn2_6/pdatata/2/pna/USO88B_COMB.seq.*
22:	/cgn2_6/pdatata/2/pna/USO88C_COMB.seq.*
23:	/cgn2_6/pdatata/2/pna/USO88D_COMB.seq.*
24:	/cgn2_6/pdatata/2/pna/USO89A_COMB.seq.*
25:	/cgn2_6/pdatata/2/pna/USO89B_COMB.seq.*
26:	/cgn2_6/pdatata/2/pna/USO89C_COMB.seq.*
27:	/cgn2_6/pdatata/2/pna/USO89D_COMB.seq.*
28:	/cgn2_6/pdatata/2/pna/USO89E_COMB.seq.*
29:	/cgn2_6/pdatata/2/pna/USO89F_COMB.seq.*
30:	/cgn2_6/pdatata/2/pna/USO89G_COMB.seq.*
31:	/cgn2_6/pdatata/2/pna/USO89H_COMB.seq.*
32:	/cgn2_6/pdatata/2/pna/USO89I_COMB.seq.*
33:	/cgn2_6/pdatata/2/pna/USO89J_COMB.seq.*
34:	/cgn2_6/pdatata/2/pna/USO90A_COMB.seq.*
35:	/cgn2_6/pdatata/2/pna/USO90C_COMB.seq.*
36:	/cgn2_6/pdatata/2/pna/USO90D_COMB.seq.*
37:	/cgn2_6/pdatata/2/pna/USO90E_COMB.seq.*
38:	/cgn2_6/pdatata/2/pna/USO91A_COMB.seq.*
39:	/cgn2_6/pdatata/2/pna/USO91B_COMB.seq.*
40:	/cgn2_6/pdatata/2/pna/USO91C_COMB.seq.*
41:	/cgn2_6/pdatata/2/pna/USO92A_COMB.seq.*
42:	/cgn2_6/pdatata/2/pna/USO92B_COMB.seq.*
43:	/cgn2_6/pdatata/2/pna/USO92C_COMB.seq.*
44:	/cgn2_6/pdatata/2/pna/USO92D_COMB.seq.*

45: /cgn2_6/prodata/2/pna/US0932D.COMB.seq.*
46: /cgn2_6/prodata/2/pna/US0922E.COMB.seq.*
47: /cgn2_6/prodata/2/pna/US0931A.COMB.seq.*
48: /cgn2_6/prodata/2/pna/US0931B.COMB.seq.*
49: /cgn2_6/prodata/2/pna/US0931C.COMB.seq.*
50: /cgn2_6/prodata/2/pna/US0931D.COMB.seq.*
51: /cgn2_6/prodata/2/pna/US0931E.COMB.seq.*
52: /cgn2_6/prodata/2/pna/US0931F.COMB.seq.*
53: /cgn2_6/prodata/2/pna/US0941A.COMB.seq.*
54: /cgn2_6/prodata/2/pna/US0941B.COMB.seq.*
55: /cgn2_6/prodata/2/pna/US6000.COMB.seq.*
56: /cgn2_6/prodata/2/pna/US6001A.COMB.seq.*
57: /cgn2_6/prodata/2/pna/US6001B.COMB.seq.*
58: /cgn2_6/prodata/2/pna/US6002A.COMB.seq.*
59: /cgn2_6/prodata/2/pna/US6002B.COMB.seq.*
60: /cgn2_6/prodata/2/pna/US6002C.COMB.seq.*
61: /cgn2_6/prodata/2/pna/US6003A.COMB.seq.*
62: /cgn2_6/prodata/2/pna/US6003B.COMB.seq.*
63: /cgn2_6/prodata/2/pna/US6004A.COMB.seq.*
64: /cgn2_6/prodata/2/pna/US6004B.COMB.seq.*
65: /cgn2_6/prodata/2/pna/US6004C.COMB.seq.*
66: /cgn2_6/prodata/2/pna/US6005.COMB.seq.*
67: /cgn2_6/prodata/2/pna/US6006.COMB.seq.*
68: /cgn2_6/prodata/2/pna/US6007.COMB.seq.*
69: /cgn2_6/prodata/2/pna/US6008A.COMB.seq.*
70: /cgn2_6/prodata/2/pna/US6008B.COMB.seq.*
71: /cgn2_6/prodata/2/pna/US6008C.COMB.seq.*
72: /cgn2_6/prodata/2/pna/US6009A.COMB.seq.*
73: /cgn2_6/prodata/2/pna/US6009B.COMB.seq.*
74: /cgn2_6/prodata/2/pna/US6010A.COMB.seq.*
75: /cgn2_6/prodata/2/pna/US6010B.COMB.seq.*
76: /cgn2_6/prodata/2/pna/US6011A.COMB.seq.*
77: /cgn2_6/prodata/2/pna/US6011B.COMB.seq.*
78: /cgn2_6/prodata/2/pna/US6012A.COMB.seq.*
79: /cgn2_6/prodata/2/pna/US6013A.COMB.seq.*
80: /cgn2_6/prodata/2/pna/US6013B.COMB.seq.*
81: /cgn2_6/prodata/2/pna/US6014A.COMB.seq.*
82: /cgn2_6/prodata/2/pna/US6014B.COMB.seq.*
83: /cgn2_6/prodata/2/pna/US6014C.COMB.seq.*
84: /cgn2_6/prodata/2/pna/US6014D.COMB.seq.*
85: /cgn2_6/prodata/2/pna/US6015.COMB.seq.*
86: /cgn2_6/prodata/2/pna/US6016.COMB.seq.*
87: /cgn2_6/prodata/2/pna/US6017.COMB.seq.*
88: /cgn2_6/prodata/2/pna/US0617.COMB.seq.*
89: /cgn2_6/prodata/2/pna/US06.COMB.seq.*
90: /cgn2_6/prodata/2/pna/US07.NEW.COMB.seq.*
91: /cgn2_6/prodata/2/pna/US08.NEW.COMB.seq.*
92: /cgn2_6/prodata/2/pna/US09.NEW.COMB.seq.*
93: /cgn2_6/prodata/2/pna/US60.NEW.COMB.seq.*

Pred, No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1738	100.0	1738	48	US-09-322-557-1	Sequence 1, Appl 1043, App
2	1721.4	99.0	3411	92	US-09-528-966-1043	Sequence 16, Appl 6770, App
3	1721	99.0	3289	28	US-08-943-570-16	Sequence 8, Appl 29, Appl
4	1710.4	98.4	3443	1	PCT-US99-24331-8	Sequence 29, Appl 9, Appl
5	1710.4	98.4	3444	87	US-60-173-373-6770	Sequence 29, Appl 9, Appl
6	1655.2	95.2	2419	1	PCT-US99-14632-29	Sequence 29, Appl 9, Appl
7	1655.2	95.2	2419	1	PCT-US99-19655-9	Sequence 29, Appl 9, Appl
8	1655.2	95.2	2419	48	US-09-322-557-3	Sequence 29, Appl 9, Appl
9	1655.2	95.2	2419	48	US-09-343-749-29	Sequence 29, Appl 9, Appl
10	1477.4	85.0	1479	1	PCT-US99-14632-1	Sequence 1, Appl 963, App
11	1477.4	85.0	1479	49	US-09-343-749-1	Sequence 2, Appl 963, App
12	992.4	57.1	1077	21	US-08-807-151-2	Sequence 963, Appl 127, App
13	881.2	50.7	1001	34	US-09-016-434-963	Sequence 127, Appl
14	466.2	28.0	357	92	US-09-528-996-127	

	Query Match	24.4%:	Score 895;	DB 1;	Length 798;	
	Best Local Similarity	36.3%:	Pred. No. 3,00e-71;			
	Matches 138;	Conservative	84;	Mismatches 141;	Indels -17;	Gaps 16;
Dd	422 CKEDNFCKD-GECIPLVNLCDGEPHCKDGSDEAHCVLFNGTSSGLVOFRIOI-WH 479					
Oy	113 CMSGICEDSGSCTCINPSNMCDGSHCPGGEDENCRYLY-GP-NFI-LQYSSQRKSMH 169					
Dd	480 VACENMTTQISDVCGLLGLGTG-NSSVPFYS-TGGGPPVINTAP-NGSLILT-PSQQ 535					
Oy	170 PWCDDMNNENYGRAACRDMGYKNMFYSSOGIVDDSGSTSEMKLNTSGANDVIYKKTHSDA 229					
Dd	536 CLDESLLILCCNKSKCKKLVTQEVSPRIYSGSDSREGAMPWVALTFDDQVCASLVS 595					
Oy	230 CSSAAVYSIKRI--ACGVNLTNSMQS-RITYGBEALGAMPMOYSLHVQVHHOCGSIIIT 286					
Dd	596 RDLVLSAHCYVGRNME-PSKWKAVLGLIHMASNITSPOIETRLIDQIVPHYNKRKN 654					
Oy	287 PEMIVTAACHV-EKPLNNPMMHTAFAGILRSFMYFGA-GYQ-VKVIISHPNDSKTKN 343					
Dd	655 DIAMHLLEAKNYNDYIQPTCLPRENOVFPRGRCSLAGMALIYGOSTADVLOEADVPL 714					
Oy	344 DIALMKLQKPLTFENDLVKPVCLLPFGMLDPQQLCWTSWGAEKEEKTSVLAANAVLL 403					
Dd	715 LSNKCOOO-MPEXNIENNVCAYEAGVDSCGSDSGCPFLMCENRRMLAGVTSFGYQ 773					
Oy	404 IETRCHSRYYIDLITIPAMICAFAFLOGNINDSCGDGSGPLVYSKNNIMWLIGDTSMGSG 463					
Dd	774 CALPNRPQYARVRPETEWI 793					
Oy	464 CAKAYRGVGNVAVFTDWI 483					
RESULT	2	STANDARD:	PRT:	798 AA.		
ID	PCT-US94-00616-2					
XX	xxxxxx					
XX						
XX						
XX						
XX	Sequence 2, Application PC/TUS9400616					
XX	GENERAL INFORMATION:					
CC	APPLICANT:					
CC	TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE					
CC	NUMBER OF SEQUENCES: 33					
CC	COMPUTER READABLE FORM:					
CC	MEDIUM TYPE: Floppy disk					
CC	COMPUTER: IBM PC compatible					
CC	OPERATING SYSTEM: PC-DOS/MS-DOS					
CC	SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)					
CC	CURRENT APPLICATION DATA:					
CC	APPLICATION NUMBER: PCT/US94/00616					
CC	FILING DATE:					
CC	CLASSIFICATION:					
CC	INFORMATION FOR SEQ ID NO: 2:					
CC	SEQUENCE CHARACTERISTICS:					
CC	LENGTH: 798 amino acids					
CC	TYPE: amino acid					
CC	STRANDEDNESS: single					
CC	TOPOLOGY: linear					
CC	MOLECULE TYPE: protein					
CC	SEQUENCE 798 AA; 88937 MW; 3292434 CN;					
Db	Query Match	24.4%:	Score 895;	DB 3;	Length 798;	
Db	Best Local Similarity	36.3%:	Pred. No. 3,00e-71;			
Oy	Matches 138;	Conservative	84;	Mismatches 141;	Indels -17;	Gaps 16;
Dd	422 CKEDNFCKD-GECIPLVNLCDGEPHCKDGSDEAHCVLFNGTSSGLVOFRIOI-WH 479					
Oy	113 CMSGICEDSGSCTCINPSNMCDGSHCPGGEDENCRYLY-GP-NFI-LQYSSQRKSMH 169					
Dd	480 VACENMTTQISDVCGLLGLGTG-NSSVPFYS-TGGGPPVINTAP-NGSLILT-PSQQ 535					
Oy	170 PWCDDMNNENYGRAACRDMGYKNMFYSSOGIVDDSGSTSEMKLNTSGANDVIYKKTHSDA 229					
Dd	536 CLDESLLILCCNKSKCKKLVTQEVSPRIYSGSDSREGAMPWVALTFDDQVCASLVS 595					
Oy	230 CSSAAVYSIKRI--ACGVNLTNSMQS-RITYGBEALGAMPMOYSLHVQVHHOCGSIIIT 286					
Dd	596 RDLVLSAHCYVGRNME-PSKWKAVLGLIHMASNITSPOIETRLIDQIVPHYNKRKN 654					
Oy	287 PEMIVTAACHV-EKPLNNPMMHTAFAGILRSFMYFGA-GYQ-VKVIISHPNDSKTKN 343					
Dd	655 DIAMHLLEAKNYNDYIQPTCLPRENOVFPRGRCSLAGMALIYGOSTADVLOEADVPL 714					
Oy	344 DIALMKLQKPLTFENDLVKPVCLLPFGMLDPQQLCWTSWGAEKEEKTSVLAANAVLL 403					
Dd	715 LSNKCOOO-MPEXNIENNVCAYEAGVDSCGSDSGCPFLMCENRRMLAGVTSFGYQ 773					
Oy	404 IETRCHSRYYIDLITIPAMICAFAFLOGNINDSCGDGSGPLVYSKNNIMWLIGDTSMGSG 463					
Dd	774 CALPNRPQYARVRPETEWI 793					
Oy	464 CAKAYRGVGNVAVFTDWI 483					

Db	480	VACAEKNTQTSDVQCCLLIGTGE-NSVPTFS-TGGGPFYVNLNTAP-NGSLILT-PSQO	535		
Oy	170	PVCDDMNMENTGRACACDMGKTKNNFYSQGIYVDSGISTFMKLTNSAGNVDYKKTYSDA	229		
Db	536	CLEDSLLIQCNTYSCGKRLTYQEVSPKRYGGSDSREGANPMVVALYEDDOVCASLVS	595		
Oy	230	CSKRAVYSLRCI--ACGVNLNSSNQS-RIVGGESALPGAMPQVSLHVNHHVCGGSIIT	286		
Db	596	RDMLVSAHCHVYGNME-PSKWKAKVLGLHMASNLTSPOIETRLIDQIVINPHNKRKN	654		
Oy	287	PEWITVTAHCV-EKPLNNPMWMTFAFAGILRQSFMEFGA-GYO-VEKYSISPNYDSKTKN	343		
Db	655	DIAMHLEMKNTYDIQYICLPEENOVFPGRICSTAGMGALTYOGSTFADVLOEADVPL	744		
Oy	344	DIATMKLQKPLTFENDLVKPCVLPFGMMLOPEQLCTWISGGATEBEKKTSEVLAAKAVL	403		
Db	715	LSNEKCOOO-PEXVNTENYMCAGYEAGVDSGCGSDSGPLMCOENNRMLLAGVTSFGYQ	773		
Oy	404	IEYQRCNSRYVNDLITPAMICAFLOGNVDSCGSDSGPLVTSKNNIMWLIQDTSHGSG	465		
Db	774	CALPNRPGVYARVPRTEWI	793		
Oy	464	CAKAYRPGVGNVMTDWI	483		
RESULT	3	STANDARD;	PRT;	638	AA.
ID	US-08-681-151-3				
XX	xxxxxx				
XX					
Dt					
XX					
De					
XX					
Sequence 3, Application US/08681151					
Patent No. 5868637					
GENERAL INFORMATION:					
APPLICANT: Au-Young, Janice					
APPLICANT: Bandman, Olga					
APPLICANT: Braxton, Scott Michael					
APPLICANT: Goll, Surya					
TITLE OF INVENTION: A NOVEL HUMAN KALLIKREIN					
NUMBER OF SEQUENCES: 4					
CORRESPONDENCE ADDRESS:					
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.					
STREET: 3174 Porter Drive					
CITY: Palo Alto					
STATE: CA					
COUNTRY: US					
ZIP: 94304					
COMPUTER READABLE FORM:					
MEDIUM TYPE: Diskette					
COMPUTER: IBM Compatible					
OPERATING SYSTEM: DOS					
SOFTWARE: FASTSEQ Version 1.5					
CURRENT APPLICATION DATA:					
APPLICATION NUMBER: US/08/681,151					
FILING DATE: Herewith					
CLASSIFICATION: 435					
PRIOR APPLICATION DATA:					
APPLICATION NUMBER:					
FILING DATE:					
ATTORNEY/AGENT INFORMATION:					
NAME: Billings, Lucy J.					
REGISTRATION NUMBER: 36,749					
REFERENCE/DOCKET NUMBER: PF-007405					
TELECOMMUNICATION INFORMATION:					
TELEPHONE: 415-855-0555					
TELEFAX: 415-845-4166					
TELEX:					
INFORMATION FOR SEQ ID NO: 3:					
SEQUENCE CHARACTERISTICS:					
LENGTH: 638 amino acids					

|||||
Db 1141 aatgaacattcgctgataagctgcagaaagcctcctgactttcaacgacctagtgaaacca 1200
Qy 1201 gtgtgtcgcgcacaccccgagatgtgtctgcagccagaaacgctctgtctgttcgcgg 1260
Db 1201 gtgtgtcgcgcacaccccgagatgtgtctgcagccagaaacgctctgtctgttcgcgg 1260
Qy 1261 tggggggcgacagagagaaagggagagaccctcagagatgctctaaagctgcgcgaagtgctc 1320
Db 1261 tggggggcgacagagagaaagggagagaccctcagagatgctctaaagctgcgcgaagtgctc 1320
Qy 1321 ctcatctgagacacagagatgcacaacagatagtctatagaaacctgataccacagcc 1380
Db 1321 ctcatctgagacacagagatgcacaacagatagtctatagaaacctgataccacagcc 1380
Qy 1381 ntgtatctgtccggtctctctgcagggagagctcgtatctctgcaggggtgcaggttgagg 1440
Db 1381 ntgtatctgtccggtctctctgcagggagagctcgtatctctgcaggggtgcaggttgagg 1440
Qy 1441 cctctgtgctcacttcgaaagacacatactcgtgtgcctgataaggagatacaagctgggtctc 1500
Db 1441 cctctgtgctcacttcgaaagacacatactcgtgtgcctgataaggagatacaagctgggtctc 1500
Qy 1501 ggtgtgcacaaagcttaacagaccagagatgtaacgggaatgtatgtattcaacggacctgg 1560
Db 1501 ggtgtgcacaaagcttaacagaccagagatgtaacgggaatgtatgtattcaacggacctgg 1560
Qy 1561 attatcgcacaaatgagggcagacggtcaatcacatggtctcgtccttgacgtcgtctt 1620
Db 1561 attatcgcacaaatgagggcagacggtcaatcacatggtctcgtccttgacgtcgtctt 1620
Qy 1621 tacaagaaacacatgagggctgttctgtctcccgctgcacgtatcttaaccttgagatgat 1680
Db 1621 tacaagaaacacatgagggctgttctgtctcccgctgcacgtatcttaaccttgagatgat 1680
Qy 1681 tcaagagctcacttcatttcaataacagtgactgtctgcgcacaaaaaataaaaaa 1738
Db 1681 tcaagagctcacttcatttcaataacagtgactgtctgcgcacaaaaaataaaaaa 1738

RESULT 2
US-09-526-996-1043
: Sequence 1043, Application US/09526996
: GENERAL INFORMATION:
: APPLICANT: Shyjan, Andrew W.
: TITLE OF INVENTION: NOCLEFT ACID MOLECULES DERIVED FROM
: FILE REFERENCE: 1600.1088-001
: CURRENT APPLICATION NUMBER: US/09/526,996
: EARLIER FILING DATE: 2000-03-15
: EARLIER APPLICATION NUMBER: 60/125,469
: EARLIER FILING DATE: 1999-03-22
: EARLIER APPLICATION NUMBER: 60/125,560
: NUMBER OF SEQ ID NOS: 1296
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 1043
: LENGTH: 3411
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-526-996-1043

Query Match 99.0%; Score 1721.4; DB 92; Length 3411;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1722; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ggcggagcgagagcgagggcgagggcgagggcgccgcttgagcgcgcaagtcac 60
Db 28 ggcggagcgagagcgagggcgagggcgagggcgccgcttgagcgcgcaagtcac 87
Qy 61 attgaacattcgaatacatcatcactcagatgtctgtatgaacaagcaagatgctctg 120

Db 88 attgaacattcgaatacatcatcactcagatgtctgtatgaacaagcaagatgctctg 147
Qy 121 aactaagggtcacacacacagctatttgacattactatgaaaaacatagataacacaggaa 180
Db 148 aactaagggtcacacacacagctatttgacattactatgaaaaacatagataacacaggaa 207
Qy 161 aacctatcccgacagacccacatgctgttcccaactgtctcagagtgcatccggctcag 240
Db 208 aacctatcccgacagacccacatgctgttcccaactgtctcagagtgcatccggctcag 267
Qy 241 tactaccgcttcccgctgcacagatgacgcccagaggtcccgacgcaagcttcacacccc 300
Db 268 tactaccgcttcccgctgcacagatgacgcccagaggtcccgacgcaagcttcacacccc 327
Qy 301 gtctgtcgcacagacccacacatcccatccggagacagtgctcacacctcaaaagataa 360
Db 328 gtctgtcgcacagacccacacatcccatccggagacagtgctcacacctcaaaagataa 387
Qy 361 gcaactgtcatcaaccttgaaacctcggggaacctcctcgttggagagctgcgtgcgtgc 420
Db 388 gcaactgtcatcaaccttgaaacctcggggaacctcctcgttggagagctgcgtgcgtgc 447
Qy 421 ctactctggaagtctatgtggagacagagctcacaactcttgagataagtgacacacctca 480
Db 448 ctactctggaagtctatgtggagacagagctcacaactcttgagataagtgacacacctca 507
Qy 481 ggtacctgcatcaacccctctaacctgtgtgataggctgcctacacgtcccgcgaggag 540
Db 508 ggtacctgcatcaacccctctaacctgtgtgataggctgcctacacgtcccgcgaggag 567
Qy 541 gacggaatcgggtgtcttcgcctcctacaggaacaaactcaatcctcaaggctgatactatc 600
Db 568 gacggaatcgggtgtcttcgcctcctacaggaacaaactcaatcctcaaggctgatactatc 627
Qy 601 cagaggaagctcctgtgcacacctgtgtgcagagcagctgtgaaagaaactaacggggcg 660
Db 628 cagaggaagctcctgtgcacacctgtgtgcagagcagctgtgaaagaaactaacggggcg 687
Qy 661 gctctcagaggacatggcctatagaataatttctctcagaacagagatagttgatgac 720
Db 688 gctctcagaggacatggcctatagaataatttctctcagaacagagatagttgatgac 747
Qy 721 agcgatccacacagctttctgaaactgaaacacaaagtgcgcggaatgctgatactataa 780
Db 748 agcgatccacacagctttctgaaactgaaacacaaagtgcgcggaatgctgatactataa 807
Qy 781 aaactgtacacagatgtagctgtctctccaagaagtggttctcttcgcgtgataacc 840
Db 808 aaactgtacacagatgtagctgtctctccaagaagtggttctcttcgcgtgataacc 867
Qy 841 tgcgggggtcaacttgaaacctcaagcgccgacagagatgtgtggcgagagcgcgctc 900
Db 868 tgcgggggtcaacttgaaacctcaagcgccgacagagatgtgtggcgagagcgcgctc 927
Qy 901 ccgggggacctgtgccttgcagagctcagcctgtcagctgcagaaagtcacagtgctgcgaag 960
Db 928 ccgggggacctgtgccttgcagagctcagcctgtcagcctgcagaaagtcacagtgctgcgaag 987
Qy 961 tccatcatcaaccccgagatgtagctgtgacacgcccacatcgctgtgaaaaaacctttaa 1020
Db 988 tccatcatcaaccccgagatgtagctgtgacacgcccacatcgctgtgaaaaaacctttaa 1047
Qy 1021 aatccatggcatgtgacggcatcttcggggatcttgagaaacacttctcatgtctcattga 1080
Db 1048 aatccatggcatgtgacggcatcttcggggatcttgagaaacacttctcatgtctcattga 1107
Qy 1081 ggcggataccaagtagaaagatgtattctcatccaatlatagactccaagaccaaagac 1140
Db 1108 ggcggataccaagtagaaagatgtattctcatccaatlatagactccaagaccaaagac 1167
Qy 1141 aatgaacattgcgtatgtagagctgcagaaagcctcgtactttcaagacctgtgtgaacca 1200
Db 1168 aatgaacattgcgtatgtagagctgcagaaagcctcgtactttcaagacctgtgtgaacca 1227

CC	REFERENCE/DOCKET NUMBER:	PF-0074US
CC	TELECOMMUNICATION INFORMATION:	
CC	TELEPHONE:	415-855-0555
CC	TELEFAX:	415-845-4166
CC	TELEX:	
CC	INFORMATION FOR SEQ ID NO: 1:	
CC	SEQUENCE CHARACTERISTICS:	
CC	LENGTH:	356 amino acids
CC	TYPE:	amino acid
CC	STRANDEDNESS:	single
CC	TOPOLOGY:	linear
CC	MOLECULE TYPE:	peptide
CC	IMMEDIATE SOURCE:	
CC	LIBRARY:	HEARNOT01
CC	CLONE:	307474
SQ	SEQUENCE	356 AA; 39995 MW; 717455 CN;
Ds	Query Match	18.8%; Score 691; DB 2; Length 356;
Oy	Best Local Similarity	35.8%; Pred. No. 4,67e-52;
Matches	118; Conservative	67; Mismatches 128; Indels 17; Gaps 16;
Ds	25 HHVACDGMQEILLSQLACKOMGLGEPSYTKL-IQEQEKEPRLTLHSNMESLNGTLLHELL	83
Oy	169 HPVCODDNWENYGRACRDMGKNMFYSOGIVDDSGSTSPMKINTS-AG-N-VDIYKKY	225
Ds	84 VNGSGCESRKISLTCTKODCGXRPAAAMNKRILIGRTSRGRRPMQCSDLOSEPSHICG	143
Oy	226 -HSDGCSSKAVAYSLPCI--ACGVNLNSSROSGRIYGEGESALPGAMPVOYSLHYQ-NVHVCG	281
Ds	144 CVLLAKKRVLVVAHQEFGRENAAYKKVVLGINLDHPSEFMQTRF-VKTIHLHRYSRAY	202
Oy	282 GSIIPEPIVTAHCVEPRPLNNPWHMTAF-GILROSMEFYGAGIOYEKVISHPYDSKT	340
Ds	203 VDYSISIELSEDISETYVRPVCLPNBPQWLIEDPTYCYITGWMGNK-MPR-LQGE	260
Oy	341 KNNIDALKLKLPFLFNOLVPCLPNPQMMLQEPQLCWISGWATTEKGKTSYLNAK	400
Ds	261 VRISLSEIKQS-Y-FDKMTTITRMICAGIESGYVDSCMGDSGGPLYCKPGCRNTLGELT	318
Oy	401 VLLIETQCNRSRYVD-NLITPMICAGIFLOGNVDSGCGDGGLYTSKNIMW-LIGDT	458
Ds	319 SWSVCSFSKVLGPQYNSVSFEVMIRKOI	348
Oy	459 SWSGSC-AKAYRPGYGNVFTDMITYROM	487
RESULT	6	STANDARD: PRT; 855 AA.
ID	US-09-027-337-2	
XX	xxxxxx	
Ds	Sequence 2, Application US/09027337B	
XX	Sequence 2, Application US/09027337B	
CC	Patent No. 5972616	
CC	GENERAL INFORMATION:	
CC	APPLICANT: O'Brien, Timothy J.	
CC	APPLICANT: Tamoto, Hirotochi	
CC	TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed In	
CC	TITLE OF INVENTION: Breast and Ovarian Carcinomas	
CC	FILE REFERENCE: D6064	
CC	CURRENT APPLICATION NUMBER: US/09/027,337B	
CC	CURRENT FILING DATE: 1998-02-20	
CC	NUMBER OF SEQ ID NOS: 13	
CC	SEQ ID NO 2	
CC	LENGTH: 855	
CC	TYPE: PRT	
CC	ORGANISM: Homo sapiens	
CC	FEATURE:	
CC	OTHER INFORMATION: Amino acid sequence of TADG-15 encoded by nucleotides	
CC	OTHER INFORMATION: 23 to 2589 of Sequence 1	

```
CC Patent No. 5972616  
SQ SEQUENCE 855 AA; 94700 MW; 3712579 CN;  
  
Query Match 18.8%; Score 632; DB 2; Length 855;  
Best Local Similarity 40.8%; Pred. No. 3,76e-32; Indels 11; Gaps 9;  
Matches 102; Conservative 58; Mismatches 79;  
  
Cc 602 CDCGRSPTROARVVGIDADGEPMQVSJHALGOGHICASLISPNMLVSAHCYIDD 661  
Cy | : ::::| : | : ||||| : : | : | : | :  
Db 243 CGVNINS-SRQSRIYGGSALPGAMPVOYSIH-VONAVCGGSTITPEMYTAAH-CVE- 298  
Oy KPL-NNWMHTATFAGILRGSEMFY-GA-GYOEKVISHPHYDSTKNNDIALMKLKPL 354  
Db 722 EYSSWVRPCLPDASHFPACKAIWTGTGMGTGYGGGALLQGKEIVNIQTGCEN-L 779  
Oy 355 TFENDLVKPCVCLEPNQMLOLPOLCWSMGATEERKTSYLNAKLVLIIETORCNRYV 414  
Db LPQQTPPRMOCVFSLSGSDSCGDSPGLSVSEADGRIFQGVYVMWGDCGAQRNKPKCV 839  
Oy 415 YDNLTTPMTIACGFLOGNVDSGCCDGSGPLVTSKNNTH-WLIQDTSMGSGCAKKAPREV 473  
Db 840 TRLPFLFRDMI 849  
Oy : : ||:  
Oy 474 GNMMFTDMT 483  
  
RESULT 7 STANDARD; PRT; 232 AA.  
ID US-08-508-448C-19  
XX xxxxxx  
DE Sequence 19, Application US/08508448C  
XX DE  
DX  
CX General Information:  
CY APPLICANT: Kazuyoshi YAMAOKA et al.  
CC TITLE OF INVENTION: NUCLEOTIC ACID SEQUENCE ENCODING  
CC TITLE OF INVENTION: TRYPSIN-LIKE ENZYME AND PROCESS FOR PRODUCING THE ENZYME  
CC NUMBER OF SEQUENCES: 25  
CC CORRESPONDENCE ADDRESS:  
CC ADDRESSEE: Wenderoth, Lind & Ponack  
CC STREET: 805 Fifteenth Street, N.W., #700  
CC City: Washington  
CC STATE: D.C.  
CC COUNTRY: U.S.A.  
CC ZIP: 20005  
CC COMPUTER READABLE FORM:  
CC MEDIUM TYPE: Diskette, 3.5 Inch, 1.44 mb  
CC COMPUTER: IBM Compatible  
CC OPERATING SYSTEM: MS-DOS  
CC SOFTWARE: Wordperfect 5.1  
CC CURRENT APPLICATION DATA:  
CC APPLICATION NUMBER: US/08/508,448C  
CC FILING DATE: July 28, 1995  
CC CLASSIFICATION: 435  
CC PRIOR APPLICATION DATA:  
CC APPLICATION NUMBER:  
CC FILING DATE:  
CC ATTORNEY/AGENT INFORMATION:  
CC NAME: Warren M. Cheek, Jr.  
CC REGISTRATION NUMBER: 33,367  
CC REFERENCE/DOCKET NUMBER:  
CC TELECOMMUNICATION INFORMATION:  
CC TELEPHONE: 202-371-8850  
CC TELEFAX:  
CC TELEX:  
CC INFORMATION FOR SEQ ID NO: 19:  
CC SEQUENCE CHARACTERISTICS:
```



```
|||||
Db 818 aaactgacacagatgacgtctgtcttccaagaagctgttcttaagcgtataagcc 877
Qy 841 tgcgggggccaacttgaactcaagccgcagagacagatgtgtggggcgagagcgcgcc 900
Db 878 tgcgggggtcaacttgaactcaagccgcagagacagatgtgtggggcgagagcgcgcc 937
Qy 901 ccgggggagcgtgagccct-gcgaggtacgcctgcagcgtccagagctccagcgtctgcgag 959
Db 938 ccgggggagcgtgagccctgcagcgtccagcgtccagagctccagcgtctgcgag 997
Qy 960 ctccatcaaccaccccgagtgagatcgtgacagccgcacacgtcgtggaagaaaccttca 1019
Db 998 ctccatcaaccaccccgagtgagatcgtgacagccgcacacgtcgtggaagaaaccttca 1057
Qy 1020 caatccatgagcatgtgagcgcacgttcgaggggagttttgagacaacttttcatgtctatg 1079
Db 1058 caatccatgagcatgtgagcgcacgttcgaggggagttttgagacaacttttcatgtctatg 1117
Qy 1080 agccggaatacgaagtagaagaaagtgatctctcatccaataatgactccaagaaccaagaa 1139
Db 1118 agccggaatacgaagtagaagaaagtgatctctcatccaataatgactccaagaaccaagaa 1177
Qy 1140 caatgacatctgcgtctgataagcgtgcagaaagccctctgacattccaagacatagtgaaac 1199
Db 1178 caatgacatctgcgtctgataagcgtgcagaaagccctctgacattccaagacatagtgaaac 1237
Qy 1200 agtgtgtctgcacaaaccagacatgtctgcagccagaaacagcctcgtctgataatccg 1259
Db 1238 agtgtgtctgcacaaaccagacatgtctgcagccagaaacagcctcgtctgataatccg 1297
Qy 1260 gtcgggggagcagcagagagagagagagagagagagagagagagagagagagagagagag 1319
Db 1298 gtcgggggagcagcagagagagagagagagagagagagagagagagagagagagagagag 1357
Qy 1320 tctcatctgagacacagagagatgcaacagcagatattctctgacaacacctgataccacagc 1379
Db 1358 tctcatctgagacacagagagatgcaacagcagatattctctgacaacacctgataccacagc 1417
Qy 1380 catgactctgctgcgagctctcctgcagagggagagctgactctcagaggtgcagctggagag 1439
Db 1418 catgactctgctgcgagctctcctgcagagggagagctgactctcagaggtgcagctggagag 1477
Qy 1440 ggcctctgctgactctgagagagaaataatactgctgctgagagggagagagagagagagag 1499
Db 1478 ggcctctgctgactctgagagagaaataatactgctgctgagagggagagagagagagagag 1537
Qy 1500 tgcgtctgcacaaagcttcaacagacagagatgtaacggaatgtgatactcaacgagactg 1559
Db 1538 tgcgtctgcacaaagcttcaacagacagagatgtaacggaatgtgatactcaacgagactg 1597
Qy 1560 gattatctgacaaatgagggagcagacgcttaatacaatgcttctgctcttcaagctcgtt 1619
Db 1598 gattatctgacaaatgagggagcagacgcttaatacaatgcttctgctcttcaagctcgtt 1657
Qy 1620 ttacagaagaaacaaatgagggagctgtgtttgtctcccggtgatacttaacttaagagatga 1679
Db 1658 ttacagaagaaacaaatgagggagctgtgtttgtctcccggtgatacttaacttaagagatga 1717
Qy 1680 ttacagaagctactcatcttttaataacagtgtaactgtctgcg 1723
Db 1718 ttacagaagctactcatcttttaataacagtgtaactgtctgcg 1761

RESULT 6
PCT-US99-14622-29
: Sequence 29, Application PC/TUS9914622
: GENERAL INFORMATION:
: APPLICANT: Wong, Alexander K.C.
: APPLICANT: Taveligian, Sean V.
: APPLICANT: Teng, David H.-F.
: APPLICANT: Myriad Genetics, Inc.
: TITLE OF INVENTION: TMRPS2 is a Tumor Suppressor
```

```
: FILE REFERENCE: 2318-202
: CURRENT APPLICATION NUMBER: PCT/US99/14622
: CURRENT FILING DATE: 1999-07-01
: EARLIER APPLICATION NUMBER: US 60/091,044
: EARLIER FILING DATE: 1998-06-29
: NUMBER OF SEQ ID NOS: 33
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO: 29
: LENGTH: 2479
: TYPE: DNA
: ORGANISM: Homo sapiens
PCT-US99-14622-29

Query Match          95.2%; Score 1655.2; DB 1; Length 2479;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1660; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 56 gtcataattgaaacatccagataccatactatcatctcgtatgtctgtgataacagcaagatgg 115
Db 1 gtcataattgaaacatccagataccatactatcatctcgtatgtctgtgataacagcaagatgg 60
Qy 116 ctttgaactcaggttcaacacacagctatttgagccttactatgataaaccatggtatccaac 175
Db 61 ctttgaactcaggttcaacacacagctatttgagccttactatgataaaccatggtatccaac 120
Qy 176 cggaaaaaccctatccgcagacagcccaactgtgtcccaactgtctaaagaggtgataccgg 235
Db 121 cggaaaaaccctatccgcagacagcccaactgtgtcccaactgtctaaagaggtgataccgg 180
Qy 236 ctcaagtaactaccgctcccgctgcacagtaagcccgaggggttcctgaaacagagcttcca 295
Db 181 ctcaagtaactaccgctcccgctgcacagtaagcccgaggggttcctgaaacagagcttcca 240
Qy 296 accccgctgctgtcagcagcagcccaatcccaatccggagagagtggtgacacctaaagacta 355
Db 241 accccgctgctgtcagcagcagcccaatcccaatccggagagagtggtgacacctaaagacta 300
Qy 356 agaaaaagcaattgacatccacttgagaccccttgagacccctctgaggagagctgtgcgtgcg 415
Db 301 agaaaaagcaattgacatccacttgagaccccttgagacccctctgaggagagctgtgcgtgcg 360
Qy 416 ctgagctactctggaagttcattcagcagcaagtgctccaaaccttgagataagatgagact 475
Db 361 ctgagctactctggaagttcattcagcagcaagtgctccaaaccttgagataagatgagact 420
Qy 476 cctcaggtactctgcatcaaacccctctaactgtgtgatagtgcgtgtcaacatgcgccggcg 535
Db 421 cctcaggtactctgcatcaaacccctctaactgtgtgatagtgcgtgtcaacatgcgccggcg 480
Qy 536 gggagagacagaaatcgggtgtgttgcctcttaagacaaacttaattccttaaggtgact 595
Db 481 gggagagacagaaatcgggtgtgttgcctcttaagacaaacttaattccttaaggtgact 540
Qy 596 catctcagaagagatcctctgacccctgtgtgcagaagacagctggaacagaaactcagagc 655
Db 541 catctcagaagagatcctctgacccctgtgtgcagaagacagctggaacagaaactcagagc 600
Qy 656 gggcgagcctgcagggagcattgaggtctaaagaataattttactctagccaagaataagtg 715
Db 601 gggcgagcctgcagggagcattgaggtctaaagaataattttactctagccaagaataagtg 660
Qy 716 atgacagcgagatccacagcttatagaataactaaacaaagtgccggcaatgtcgatatct 775
Db 661 atgacagcgagatccacagcttatagaataactaaacaaagtgccggcaatgtcgatatct 720
Qy 776 ataaaaaacctgtacacacagtgatgcctgtcttccaagaacagtggtttcttaacgtcta 835
Db 721 ataaaaaacctgtacacacagtgatgcctgtcttccaagaacagtggtttcttaacgtcta 780
Qy 836 tagcctcgagggttcaacttgaactcaagcgcagagcagagatctgtggcgagcagagcg 895
Db 781 tagcctcgagggttcaacttgaactcaagcgcagagcagagatctgtggcgagcagagcg 840
```



```

Oy 956 gaggtccatcatcaccacccgagtgatcgtagacagccgcccactgctgtgaaaaactc 1015
    |||||||
Db 901 gaggtccatcatcaccacccgagtgatcgtagacagccgcccactgctgtgaaaaactc 960
Oy 1016 ttaacaatcatatgcatatgtagacgcatctgcgggagatttgagacaactcttcattct 1075
    |||||||
Db 961 ttaacaatcatatgcatatgtagacgcatctgcgggagatttgagacaactcttcattct 1020
Oy 1076 atgtagcccgagataccaagttagaaaaagtgtattctcattccaaatatatgattccaagaca 1135
    |||||||
Db 1021 atgtagcccgagataccaagttagaaaaagtgtattctcattccaaatatatgattccaagaca 1080
Oy 1136 agaaacatgacatctgcgctgtagatgagctgcagaagcctctgactcttcaacgacactgta 1195
    |||||||
Db 1081 agaaacatgacatctgcgctgtagatgagctgcagaagcctctgactcttcaacgacactgta 1140
Oy 1196 aaccagtgctctgcgcccacccagcatgattgctgcagccagaacagctctgctgattt 1255
    |||||||
Db 1141 aaccagtgctctgcgcccacccagcatgattgctgcagccagaacagctctgctgattt 1200
Oy 1256 ccgggtgaggggccacccagagagaaaggaaagaaacctcagaagtctgtaacgctgcgaag 1315
    |||||||
Db 1201 ccgggtgaggggccacccagagagaaaggaaagaaacctcagaagtctgtaacgctgcgaag 1260
Oy 1316 tgccttcattatgagacacagagatgcaacagagataatgctatgacaacactgatacac 1375
    |||||||
Db 1261 tgccttcattatgagacacagagatgcaacagagataatgctatgacaacactgatacac 1320
Oy- 1376 cagccatgacatctgctgcgctcctgcagaggagagagctgactcttgcagaggtagacatg 1435
    |||||||
Db 1321 cagccatgacatctgctgcgctcctgcagaggagagagctgactcttgcagaggtagacatg 1380
Oy 1436 gagggcctctgctgcaatctcgaagaacaatactgctgctgtaaggagagataaaactgag 1495
    |||||||
Db 1381 gagggcctctgctgcaatctcgaagaacaatactgctgctgtaaggagagataaaactgag 1440
Oy 1496 gtccgcgctgctgcacaaactcagacacagagagtagcggagatgtagatgataccag 1555
    |||||||
Db 1441 gtccgcgctgctgcacaaactcagacacagagagtagcggagatgtagatgataccag 1500
Oy 1556 actggaattatcgacaataatgagggagagagcgtataatccacatgctcttcgtcctgagct 1615
    |||||||
Db 1501 actggaattatcgacaataatgagggagagagcgtataatccacatgctcttcgtcctgagct 1560
Oy 1616 cgtttacaagaagaacaatgagggctggttctgctcccgctcagatgattactcttaagg 1675
    |||||||
Db 1561 cgtttacaagaagaacaatgagggctggttctgctcccgctcagatgattactcttaagg 1620
Oy 1676 atgattcagaggtcactcattcttataaagaatggaactgtctggc 1723
    |||||||
Db 1621 atgattcagaggtcactcattcttataaagaatggaactgtctggc 1668

```

RESULT 8

```

US-09-323-597-3
; Sequence 3, Application US/09323597
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel E
; APPLICANT: Hubert, Rene S
; APPLICANT: Leong, Kahan
; APPLICANT: Raitano, Arthur B
; APPLICANT: Saifan, Douglas C
; TITLE OF INVENTION: NOVEL TUMOR ANTIGEN USEFUL IN DIAGNOSIS AND THERAPY OF
; FILE REFERENCE: 1703-007, US1
; CURRENT APPLICATION NUMBER: US/09/323,597
; EARLIER FILING DATE: 1998-06-01
; EARLIER APPLICATION NUMBER: 60/087,598
; EARLIER FILING DATE: 1998-06-01
; EARLIER APPLICATION NUMBER: 60/091,474
; EARLIER FILING DATE: 1998-06-29
; EARLIER APPLICATION NUMBER: 60/129,521

```

```

; EARLIER FILING DATE: 1999-04-14
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patentl Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2479
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; JOURNAL: Genomics
; VOLUME: 44
; PAGES: 309-320
; DATE: 1997
US-09-323-597-3

```

```

Query Match 95.2% Score 1655.2 DB 48: Length 2479;
Best Local Similarity 99.5% Pred. NO. 0;
Matches 1660; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

```

Oy 56 gtcatattgaacattccagatatacctatcatctactcgtatgctgtgtatatacagcaagatg 115
    |||||||
Db 1 gtcatattgaacattccagatatacctatcatctactcgtatgctgtgtatatacagcaagatg 60
Oy 116 ctgtgaactcaggggtcacacacagcatattggaaccttactatgaaacaacatgatacaac 175
    |||||||
Db 61 ctgtgaactcaggggtcacacacagcatattggaaccttactatgaaacaacatgatacaac 120
Oy 176 cggaaaaacccctatccgcgacacagccactgtgtgtcccaactgtctacagagtgtacccg 235
    |||||||
Db 121 cggaaaaacccctatccgcgacacagccactgtgtgtcccaactgtctacagagtgtacccg 180
Oy 236 ctcaactactacccgtcccccgtgcccagtagcccccagaggtctctgacgacaggttcca 295
    |||||||
Db 181 ctcaactactacccgtcccccgtgcccagtagcccccagaggtctctgacgacaggttcca 240
Oy 296 accccgtctctgcacgacagccaaatcccaatcccggaacaggtgtgcacctcaaaagacta 355
    |||||||
Db 241 accccgtctctgcacgacagccaaatcccaatcccggaacaggtgtgcacctcaaaagacta 300
Oy 356 agaaagacactgtgcatacctctgacccctgggagacctctcgttggagactgctgcggcg 415
    |||||||
Db 301 agaaagacactgtgcatacctctgacccctgggagacctctcgttggagactgctgcggcg 360
Oy 416 ctgacactcctggaagtcatagggcagcaagtgctccaaactctggtgtaagtgagct 475
    |||||||
Db 361 ctgacactcctggaagtcatagggcagcaagtgctccaaactctggtgtaagtgagct 420
Oy 476 cctcagtaactgcatcaaaccccttaactgtgtgtagtgcgtgtcacactgcccggcg 535
    |||||||
Db 421 cctcagtaactgcatcaaaccccttaactgtgtgtagtgcgtgtcacactgcccggcg 480
Oy 536 gggagagacgagaatcgtgtgtgtctgctcctcctacgagccaaactctcctcaagtgact 595
    |||||||
Db 481 gggagagacgagaatcgtgtgtgtctgctcctcctacgagccaaactctcctcaagtgact 540
Oy 596 catctcagagagatcctgagccctgtgtgtccaaagagactggaacgagaaactacggcg 655
    |||||||
Db 541 catctcagagagatcctgagccctgtgtgtccaaagagactggaacgagaaactacggcg 600
Oy 656 gggcgccctgcagagacatggtgctaagaataattttactcttagcaagaagatagtg 715
    |||||||
Db 601 gggcgccctgcagagacatggtgctaagaataattttactcttagcaagaagatagtg 660
Oy 716 atgacagcgatccacagcttatgaaactgaaacgaaagtgccggacatgtcatatct 775
    |||||||
Db 661 atgacagcgatccacagcttatgaaactgaaacgaaagtgccggacatgtcatatct 720
Oy 776 ataaaaactgtacacagagtgatgctgtcttctcaaaagcggtgttcttcaagctga 835
    |||||||
Db 721 ataaaaactgtacacagagtgatgctgtcttctcaaaagcggtgttcttcaagctga 780
Oy 836 tagcctgagggttcaacttgaactcaaaccccgacagagagagattgtggcgcgagagcg 895
    |||||||

```



```

Oy 956 gagctcaccatcccccagtgatcgtgacagccgcccactgctgtaaaaaaacc 1015
Db 901 gagctccatccatcccccagtgatcgtgacagccgcccactgctgtaaaaaaacc 960
Oy 1016 ttaacaatccatccatcccccagtgatcgtgacagccgcccactgctgtaaaaaaacc 1075
Db 961 ttaacaatccatccatcccccagtgatcgtgacagccgcccactgctgtaaaaaaacc 1020
Oy 1076 atgagccgagatcccaaatgaaagtgtatctccatcccaaatatgactccaagacca 1135
Db 1021 atgagccgagatcccaaatgaaagtgtatctccatcccaaatatgactccaagacca 1080
Oy 1136 agaacaatgacatctgcgtgatagaaagctgcagaagcctctgactccaagacctgta 1195
Db 1081 agaacaatgacatctgcgtgatagaaagctgcagaagcctctgactccaagacctgta 1140
Oy 1196 aaccagctgtctgcgcacacacacagcatgactgctgacgcaaaacagcctctgactt 1255
Db 1141 aaccagctgtctgcgcacacacacagcatgactgctgacgcaaaacagcctctgactt 1200
Oy 1286 ccgagctgagggccacacagaggaaggaagacccatcagaatgctgaaacgctgcacaag 1315
Db 1201 ccgagctgagggccacacagaggaaggaagacccatcagaatgctgaaacgctgcacaag 1260
Oy 1316 tgcctcattgagacacagagatgcaacacagagataatgctatgacaacccatgacacac 1375
Db 1261 tgcctcattgagacacagagatgcaacacagagataatgctatgacaacccatgacacac 1320
Oy 1376 cagccatgactgctgcgcgcgtctcctgcagggaggaacgtcgaatctctgcagggctgacg 1435
Db 1321 cagccatgactgctgcgcgcgtctcctgcagggaggaacgtcgaatctctgcagggctgacg 1380
Oy 1436 gagggccctgctcactccacgaagaacataatctgctgagctgaaaggatataaagctgg 1495
Db 1381 gagggccctgctcactccacgaagaacataatctgctgagctgaaaggatataaagctgg 1440
Oy 1496 gttcctgctgctgcacaaagcttaccagaccagagatgtaacgggaatgtaagtaattacc 1555
Db 1441 gttcctgctgctgcacaaagcttaccagaccagagatgtaacgggaatgtaagtaattacc 1500
Oy 1556 actggaattatccgaacaatgagagcgacagcgttaacacatggtctctgctcctgagact 1615
Db 1501 actggaattatccgaacaatgagagcgacagcgttaacacatggtctctgctcctgagact 1560
Oy 1616 cgtttacaagaagaacaatgagcggtgttctgctcccgctgcatgacttactctagaag 1675
Db 1561 cgtttacaagaagaacaatgagcggtgttctgctcccgctgcatgacttactctagaag 1620
Oy 1676 atgattcagaggtcactcattcttataaacaagtgaactgtctggc 1723
Db 1621 atgattcagaggtcactcattcttataaacaagtgaactgtctggc 1668

```

```

RESULT 10
PCT-US99-14622-1
Sequence 1, Application PC/TUS9914622
GENERAL INFORMATION:
APPLICANT: Wong, Alexander K.C.
APPLICANT: Tavliglian, Sean V.
APPLICANT: Teng, David H.-F.
APPLICANT: Myriad Genetics, Inc.
TITLE OF INVENTION: TMRPS2 is a Tumor Suppressor
FILE REFERENCE: 2318-202
CURRENT APPLICATION NUMBER: PCT/US99/14622
EARLIER FILING DATE: 1999-07-01
EARLIER APPLICATION NUMBER: US 60/091,044
NUMBER OF SEQ ID NOS: 33
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 1
LENGTH: 1479
TYPE: DNA
ORGANISM: Homo sapiens

```

```

FEATURE:
NAME/KEY: CDS
LOCATION: (1) .. (1476)
FEATURE:
NAME/KEY: conflict
LOCATION: (724)
OTHER INFORMATION: Listed as T in GenBank Accession NO. U75329
FEATURE:
NAME/KEY: conflict
LOCATION: (985)
OTHER INFORMATION: Listed as C in GenBank Accession NO. U75329
FEATURE:
NAME/KEY: conflict
LOCATION: (1466)
OTHER INFORMATION: Listed as A in GenBank Accession NO. U75329
FEATURE:
NAME/KEY: conflict
LOCATION: (1471)
OTHER INFORMATION: Listed as A in GenBank Accession NO. U75329
FEATURE:
NAME/KEY: allele
LOCATION: (478)
OTHER INFORMATION: This base can be G or A with G being the more
OTHER INFORMATION: common allele. The codon will change from Val to
OTHER INFORMATION: Met.
FEATURE:
NAME/KEY: allele
LOCATION: (777)
OTHER INFORMATION: This base can be C or T with C being the more
OTHER INFORMATION: common allele. The codon is unaffected with both
OTHER INFORMATION: alleles encoding Gly.
FEATURE:
NAME/KEY: allele
LOCATION: (768)
OTHER INFORMATION: This base can be C or T with C being the more
OTHER INFORMATION: common allele. This is a silent polymorphism.
FEATURE:
NAME/KEY: allele
LOCATION: (834)
OTHER INFORMATION: This base can be C or T with C being the more
OTHER INFORMATION: common allele. This is a silent polymorphism.
FEATURE:
NAME/KEY: allele
LOCATION: (625)
OTHER INFORMATION: This base can be T or A with T being the more
OTHER INFORMATION: common allele. The codon will change from Phe to
OTHER INFORMATION: Ile
PCT-US99-14622-1

```

```

Query Match 85.0%; Score 1477.4; DB 1; Length 1479;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Oy 112 atgcttgaactcagggtcacaccacagctatgtgacttataagaacaatgatgac 171
Db 1 atgcttgaactcagggtcacaccacagctatgtgacttataagaacaatgatgac 60
Oy 172 caaccgaaaacccctatccgcacacagccactggtgctccacactgctacagagtgat 231
Db 61 caaccgaaaacccctatccgcacacagccactggtgctccacactgctacagagtgat 120
Oy 232 ccggtcagtaactaccgctcccgctgcccagtaagcccgaggggtcctgacgagct 291
Db 121 ccggtcagtaactaccgctcccgctgcccagtaagcccgaggggtcctgacgagct 180
Oy 292 tccaaccccgctgtctgacagcagcccaaatcccatcccgagagagtgacccccaag 351
Db 181 tccaaccccgctgtctgacagcagcccaaatcccatcccgagagagtgacccccaag 240

```

DT 01-FEB-1994 (rel. 28, Created)
DT 01-FEB-1994 (rel. 28, Last sequence update)
DT 01-NOV-1997 (rel. 35, Last annotation update)
DE SERINE PROTEASE HEPsin (EC 3.4.21.-).
GN HPN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-LIVER;
RA MEDLINE; 93305733.
RA FARLEY D., REYMOND F., NICK H.;
RT "Cloning and sequence analysis of rat hepsin, a cell surface serine
RT protease."
RT Blochm. Biophys. Acta 1173:350-352(1993).
CC -1- FUNCTION: PLAYS AN ESSENTIAL ROLE IN CELL GROWTH AND MAINTENANCE
CC OF CELL MORPHOLOGY.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X70900; CAA50256.1; -
DR PIR; S32013; S32013.
DR HSSP; P00763; 1PPO.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PFAM; PF00089; trypsin; 1.
KN Hydrolyase; Serine protease; Transmembrane; Signal-anchor.
FT CHAIN 1 161 NON-CATALYTIC CHAIN (POTENTIAL).
FT DOMAIN 1 16 CATALYTIC CHAIN (POTENTIAL).
FT TRANSMEM 1 16 CYTOPLASMIC (POTENTIAL).
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT (POTENTIAL).
FT EXTRACELLULAR (POTENTIAL).
FT CATALYTIC. (POTENTIAL).
FT ACT_SITE 162 416 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 202 202 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 236 236 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 352 352 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 152 276 INTERCHAIN (BY SIMILARITY).
FT DISULFID 187 203 BY SIMILARITY.
FT DISULFID 321 337 BY SIMILARITY.
FT DISULFID 348 380 BY SIMILARITY.
FT CARBOHYD 111 111 POTENTIAL.
SQ SEQUENCE 416 AA; 44926 MW; 2FF69FB CRC32;

Query Match 17.83; Score 652; DB 1; Length 416;
Best Local Similarity 40.68; Pred. No. 1.27e-142;
Matches 101; Conservative 52; Mismatches 83; Indels 13; Gaps 9;

Oy 427 GFLQGNVDSQGDGGPLY-TSK-NNI--WMLIGDTWSGSCAKAYRPGVGNVFTDM 482
Db 399 IFOAIRKHS 407
Oy 483 IYQMRADG 491

RESULT 14
ID PLMN_MOUSE STANDARD; PRT; 812 AA.
AC P20918;
DT 01-FEB-1991 (rel. 17, Created)
DT 01-FEB-1991 (rel. 17, Last sequence update)
DT 15-JUL-1999 (rel. 38, Last annotation update)
DE PLASMINOGEN PRECURSOR (EC 3.4.21.7) (CONTAINS: ANGIOSTATIN).
GN PLG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE; 91184812.
RA DEGEN S.J., BELL S.M., SCHAEFER L.A., ELLIOTT R.M.;
RT "Characterization of the cDNA coding for mouse plasminogen and
RT localization of the gene to mouse chromosome 17.";
RT Genomics 8:49-61(1990).
RN
RN [2]
RP CHARACTERIZATION OF ANGIOSTATIN, AND PARTIAL SEQUENCE.
RX MEDLINE; 95042728.
RA O'REILLY M.S., HOLMGREN L., SHING Y., CHEN C., ROSENTHAL R.A.,
RA MOSS M., LANE W.S., CAO Y., SAGE E.H., FOLKMAN J.;
RT "Angiostatin: a novel angiogenesis inhibitor that mediates the
RT suppression of metastases by a Lewis lung carcinoma."
RT Cell 79:315-328(1994).
RL
RL -1- FUNCTION: PLASMIN DISSOLVES THE FIBRIN OF BLOOD CLOTS AND ACTS AS
CC A PROTEOLYTIC FACTOR IN A VARIETY OF OTHER PROCESSES INCLUDING
CC EMBRYONIC DEVELOPMENT, TISSUE REMODELING, TUMOR INVASION,
CC AND INFLAMMATION. IN OVULATION IT WEAKENS THE WALLS OF THE
CC GRAFVIAN FOLLICLE. IT ACTIVATES THE UROKINASE-TYPE PLASMINOGEN
CC ACTIVATOR, COLLAGENASES AND SEVERAL COMPLEMENT ZYMOGENS, SUCH
CC AS C1 AND C5. IT CLEAVES FIBRIN, FIBRONECTIN, THROMBOSPONDIN,
CC LAMININ AND VON WILLEBRAND FACTOR.
CC -1- FUNCTION: ANGIOSTATIN IS AN ANGIOGENESIS INHIBITOR THAT BLOCKS
CC NEOVASCULARIZATION AND GROWTH OF EXPERIMENTAL PRIMARY AND
CC METASTATIC TUMORS IN VIVO.
CC -1- ENZYME REGULATION: CONVERTED INTO PLASMIN BY PLASMINOGEN
CC ACTIVATORS, BOTH PLASMINOGEN AND ITS ACTIVATOR BEING BOUND TO
CC FIBRIN. CANNOT BE ACTIVATED WITH STREPTOKINASE.
CC -1- MISCELLANEOUS: PLASMIN IS INACTIVATED BY ALPHA-2-ANTIPLASMIN
CC IMMEDIATELY AFTER DISSOCIATION FROM THE CLOT.
CC -1- MISCELLANEOUS: IN THE PRESENCE OF THE INHIBITOR, THE ACTIVATION
CC INVOLVES ONLY CLEAVAGE AFTER ARG-581, RESULTING IN 2 CHAINS HELD
CC TOGETHER BY 2 DISULFIDE BONDS. WITHOUT THE INHIBITOR, THE
CC ACTIVATION INVOLVES ALSO REMOVAL OF THE ACTIVATION PEPTIDE.
CC -1- SIMILARITY: CONTAINS 5 KRINGLE REGIONS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. BELONGS TO THE PLASMINOGEN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; J04766; AAA50168.1; -
DR PIR; A38514; A38514.
DR HSSP; P00747; 1PMK.
DR MGD; MGI:97620; PLG.
DR PROSITE; PS00021; KRINGLE_1; 4.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR PROSITE; PS0070; KRINGLE_2; 5.


```

15 GCACAGGGGGGGGSSCGCCCTGGAAACCGGAGGATATGATGACATTCAGATA 74
OY cctatcattactgagatgctgctgtaataacagcaagatggcttgtaacagagtcacacac 137
78 cctatcattactgagatgctgctgtaataacagcaagatggcttgtaacagagtcacacac 137
Db cctatcattactgagatgctgctgtaataacagcaagatggcttgtaacagagtcacacac 134
OY 138 agctatggaaccttactatgataaacaatgataccaacccggaacccatccatccgaca 197
135 wgtatattgacaccttactatgataaacaatgataccaacccggaacccatccatccgaca 194
OY 198 gcccaattgctgctccacacgctgctcagagtgatcccggtcagtaactaccgctcccgct 237
195 GCCCACTGAGTCCACACGTCTACAGAGGTGCTCCGGCTCACTACTACCCCTCCCGCT 254
OY 258 gcccaattgctgctccacacgctgctcagagtgatcccggtcagtaactaccgctcccgct 317
255 GCCCACTGAGTCCACACGTCTACAGAGGTGCTCCGGCTCACTACTACCCCTCCCGCT 314
OY 318 caaatcccca--tccggaacagtgctgacactcaagaactaagaagaacactgctgacac 375
315 CAATCCCATGCTCKGAGACNTGTCACCTCAAGACTAAGAAACGACTGTCATCAC 374
OY 376 ttgaaccttggaacctctcctgctggaagctgctgctgctgctgctgctgctgctgctgct 435
375 TTGACCCCTGGGACCTTCCTGCTGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 434
OY 436 atgggagcaagaatgctcacaactctggaatagatgtagatgtagatgtagatgtagatgtagat 495
435 ATGGGAGCAAGAATGCTCACAACCTCTGGATAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 494
OY 496 cccctcaactgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 554
499 CCTCTCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 554
OY 555 tcttgcgctcctcagcaacactcactcctcagtgtaactcactcagaggaagtcctg 614
555 TTTTGCCCTCTGAGGACCAACTCTCTGAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 614
OY 615 gaaacctgctgctcagcaacactcagaggaactcagaggaactcagaggaactcagaggaact 674
615 GAAACCTGCTGCTCAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 674
OY 675 gggcctaaagaataatcttactcctcagcaaggaatagtgatgagagagagagagagagagagag 734
675 GGGCCTAAAGAAATATTTTACTCTAGCAAGAAATAGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 734
OY 735 cttctgaaactgaaacaaagtgcgagcaatgcgaatcctcctcctcctcctcctcctcctcctcct 794
735 CTTTATGAACCTGAACCAAGTGCAGCAATGCTGATCTATATAAAACTGTACACAG 794
OY 795 tgaatccgcttctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 854
795 TGAATCCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 854
OY 855 gaaactcaagcccgagagagagatgctgagcgagc-gagagcgagcctcccgagagagagagagag 913
855 GAACTCAAGCCCGCAGCAAGAGATGCTGCGCGGCAANNAGCGCGCTCCCGGGGGCTTGGC 914
OY 914 cctgagagagagagagagagagagagagagagagagagagagagagagagagagagagagagagag 973
915 CCTGAGAGAGTCAAGCTGCAAGTCCACACCTCCACACCTCCACACCTCCACACCTCCACACCTCCACAC 974
OY 974 ccgaatgagatgctgacagcgcc 996
975 CCGAATGATCTGTCACCGCC 997

```

```

RESULT 14
US-09-526-996-127/c
; Sequence 127, Application US/09526996
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.

```

```

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES DERIVED FROM
; FILE OF INVENTION: HUMAN PROSTATE AND BREAST LIBRARIES
; FILE REFERENCE: 1600.1088-001
; CURRENT APPLICATION NUMBER: US/09/526,996
; EARLIER FILING DATE: 1999-03-22
; EARLIER APPLICATION NUMBER: 60/125,469
; EARLIER FILING DATE: 1999-03-22
; EARLIER APPLICATION NUMBER: 60/125,560
; NUMBER OF SEQ ID NOS: 1296
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 127
; LENGTH: 537
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(537)
; OTHER INFORMATION: n = A,T,C or G
US-09-526-996-127

```

```

Query Match      28.0%; Score 486.2; DB 92; Length 537;
Best Local Similarity 97.4%; Pred. No. 3,1e-132;
Matches 524; Conservative 0; Mismatches 11; Indels 3; Gaps 3;

```

```

OY 578 tcatcctcagtgatgctatctcctcagagagagagagagagagagagagagagagagagagagag 637
Db 537 TCACTCTTCAAGGTGATCTATCTCAGAGAGAGTCTCGGACCTGTGTCAGAGACGAC 478
OY 638 ggaacagagagagagagagagagagagagagagagagagagagagagagagagagagagagagagag 697
638 GGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 697
OY 698 gtagcaagagagagagagagagagagagagagagagagagagagagagagagagagagagagagagag 757
698 GTAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 757
OY 758 cggagagagagagagagagagagagagagagagagagagagagagagagagagagagagagagag 816
758 CGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 816
OY 817 gtagctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 876
817 GTAGCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT 876
OY 877 attgagagagagagagagagagagagagagagagagagagagagagagagagagagagagagagag 936
877 ATTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 936
OY 937 cagaagagagagagagagagagagagagagagagagagagagagagagagagagagagagagagag 996
937 CAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 996
OY 997 cactgctgagagagagagagagagagagagagagagagagagagagagagagagagagagagagagagag 1055
997 CACTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1055
OY 1056 gagacaatcttcatgctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcct 1113
1056 GAGACAATCTTCTCATGCTCTATGAGAGCGGATACCAAGTAGAANAAGATTTNTCAT 1

```

```

RESULT 15
US-09-526-996-1044
; Sequence 1044, Application US/09526996
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES DERIVED FROM
; FILE REFERENCE: 1600.1088-001
; CURRENT APPLICATION NUMBER: US/09/526,996
; CURRENT FILING DATE: 2000-03-15

```



```

OY 342 NNDIALMKLCKPLTFNDLVAVPCLPNPGMMLQPEQLCWISGWMGTEKGTSEVLNAKY 401
DB 948 RIISLEHCOS-Y-FDMKTTTRMICAAGESGTQSCMGSDGPLYCEKPGGRWTLFGLTS 1005
OY 402 LLEIETORCNSRYVD-NLIRPAMICAGFLQGNVDSGCGSGPLVTSKNIMW-LIGDTS 459
DB 1006 WSGVCFSKVLGPGVYSNVSYFVEMIKRQI 1034
OY 460 WSGSGC-AKAYRPGVGNVWFTDMYROM 487

RESULT 2
ID 097506 PRELIMINARY; PRT; 643 AA.
AC 097506;
DT 01-MAY-1999 (TEMBLrel. 10, Created)
DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)
DE KALLIKREIN.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PLASMA.
RA TAKAHASHI T., KIMURA A., OKIMURA H., HAMABATA T.;
RT "Porcine liver plasma kallikrein.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB022425; BAA37147.1;
DR HSSP: P00766; 1CHG.
SO SEQUENCE 643 AA; 72227 MW; F0154450 CRC32;

Query Match
Best Local Similarity 42.7%; Score 751; DB 6; Length 643;
Matches 108; Conservative 48; Mismatches 89; Indels 8; Gaps 7;

DB 379 SLFLCRSGDHSACATKNTNTVGTSTSEFLDEMPQVSLQAKLRQNLGGSTIGHQWVL 438
OY 237 SLR-CIACGVNLSSR-OSRIVGESALPGAMPQVSLHVO-NV-HVCGSITTPEMIV 291
DB 439 TAACEFD-GLSLPDIMRIYGLINISEITRETPSQVKEIIHONKYLBSGDIALKL 497
OY 292 TAACEFD-GLSLPDIMRIYGLINISEITRETPSQVKEIIHONKYLBSGDIALKL 350
DB 498 ETEPLNTDFQKPICLPSRDDTNVYITNCWATGCTEKEGKQIQLKVNIPVLSNECQ 557
OY 351 QKPLETNDLVKPCPLPMPGMLQPEQLCWISGWMGTEKGTSEVLNAKYLLEIETORCN 410
DB 558 KSTR-DKIKIKOMICAGYKGGKADCKGSGGPLYCKYKNGIMHLYVTSTSGEGCARREP 616
OY 411 SRVYVDLILIPAMICAGFLQGNVDSGCGSGPLVTSKNIMW-LIGDTSWGSACAKAYRP 470
DB 617 GYTKVIEYMDWI 629
OY 471 GYTKVIEYMDWI 483

RESULT 3
ID 092319 PRELIMINARY; PRT; 1113 AA.
AC 092319;
DT 01-MAY-1999 (TEMBLrel. 10, Created)
DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)
DE LOW DENSITY LIPOPROTEIN RECEPTOR RELATED PROTEIN 4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sclurognathia; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE: 98429596.
RA TOMITA Y., KIM D.-H., MAGOORI K., FUJINO T., YAMAMOTO T.T.;
RT "A novel low-density lipoprotein receptor-related protein with type II
membrane protein-like structure is abundant in heart.";

```

```

RL EMBL: 124:784-789(1998).
DR EMBL: AB013874; BAA34371.1; -.
DR HSSP: P00763; IDPO.
DR PROSITE: PS01209; LDLRA_1; 6.
KM Glycoprotein.
SO SEQUENCE 1113 AA; 122984 MW; ABAD31E CRC32;

Query Match
Best Local Similarity 39.1%; Score 702; DB 4; Length 283;
Matches 97; Conservative 50; Mismatches 91; Indels 6; Gaps 5;

DB 723 CODNELECANHE-CVPRDLCDGWDGSDSDSDGCVTLTKNGNSSSLTLVHRKA-KE-H 779
OY 113 CSNSGIECDSSGTCINPNSMCDGVSHCPGEGEDENRCVRL-YG-PNFIQVYSQKSMH 169
DB 780 HVCADGWRRLTSQLACKOMKGLGEPVTKL-IPQEGQO-WLRLYPMENLNGSTLQELLV 837
OY 170 PCQDDNMENYGRACRMDGKKNMFYSSQGIYVDSGSTSPKMLNTS-AG-NVDYKRY-- 225
DB 838 YRHSCPSRSEISLCSKQDCGRPARAAMKRIIGKTRSPGRMPQCSLQSPSGHICCG 897
OY 226 HSDACSKAVVSLRCI--ACGVNLSSRSRIVGGSALPGAMPQVSLHVO-NVHVC 282
DB 898 VLIAKKWLTVACFCGREDADYMK-VFQINNLDPSGPMOTRF-VKTIILPYSRAV 955
OY 283 SITTPETVTAACVE-KPLNPMHTARA-GILRSFMYGAGYEVKVISHPNDSTK 340
DB 956 VVDIISVELSDINETSYPVCLPSPEEYLDPTCYITGNGHGNK-MPEK-LQEGE 1013
OY 341 KNNDIALMKLQKPLTFNDLVAVPCLPNPGMMLQPEQLCWISGWMGTEKGTSEVLNAKY 400
DB 1014 VRIIPEQCS-Y-FDMKTTTRMICAAGESGTQSCMGSDGPLYCEKPGGRWTLFGLT 1071
OY 401 VLIETORCNSRYVD-NLIRPAMICAGFLQGNVDSGCGSGPLVTSK-NIMW-LIGDT 458
DB 1072 WSGVCFSKVLGPGVYSNVSYFVEMIKRQI 1101
OY 459 WSGSGC-AKAYRPGVGNVWFTDMYROM 487

RESULT 4
ID 095518 PRELIMINARY; PRT; 283 AA.
AC 095518;
DT 01-MAY-1999 (TEMBLrel. 10, Created)
DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)
DE DL117OK4.2 (NOVEL TRYPSIN FAMILY PROTEIN WITH CLASS A LDL RECEPTOR
DOMAINS) (FRAGMENT).
GN DL117OK4.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA BLAKEY S.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL022314; CAI18442.1; -.
DR HSSP: P00763; IDPO.
KM Receptor.
FT NON TER
SO SEQUENCE 283 AA; 31292 MW; E0FE50DB CRC32;

Query Match
Best Local Similarity 39.1%; Score 702; DB 4; Length 283;
Matches 97; Conservative 50; Mismatches 91; Indels 6; Gaps 5;

DB 38 CDGGLGSPS-SRIYGAVSSEGEPMQASLQVGRHICGALLADRMVITPAHCFQEDSM 96
OY 243 GCVNLSSRSRIVGGSALPGAMPQVSLHVO-NVHVCSSITTPETVTAACVE-EKPL 301
DB 97 ASYVLMTVFQKRWONSRRPGEVSKVSRLLRPHYEEDSHDYVALLDLHPVNSAAV 156
OY 302 NNPWHMTAFAGILRSFMYGAG-YEVKVISHPNDSTKKNNDIALMKLQKPLTFNDLV 360

```

Db 2285 TGGCTCTGGCGGCGTGCATTTGGATATCATGTGCACCTGCGCAATGCGCCAGG 2344
Qy 1528 gtgtacgggaatgtgagtatttcacggactgcat 1562
Db 2345 GTGTATGCCCGGCTCCCAAGGTTCAACAGATGTGAT 2379

RESULT 3
PCT-US94-00616-1
Sequence 1, Application PC/TUS9400616
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE
NUMBER OF SEQUENCES: 33
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/00616
FILING DATE:
CLASSIFICATION:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2581 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
PCT-US94-00616-1

Query Match 6.7%; Score 115.8; DB 6; Length 2581;
Best Local Similarity 50.9%; Pred. No. 3.4e-22;
Matches 354; Conservative 0; Mismatches 332; Indels 9; Gaps 3;

Qy 874 aggattgtggcgcgagagcgcgctcccggggctggcctggcctgcaagtcagctgac 933
Db 1688 AAGATTGTGCGAGAGAGTACTCCAGAGAGAGAGCGCTGGCTGGCTGTGAT 1747
Qy 934 gtccagagacgctccagctgtgggagagctccatccaccgagctgagctgacagcc 993
Db 1748 TTGACGCTAACAGGCTGTGGAGCTTCTGTGACAGAGGATGGCTGTGCGGCC 1807
Qy 994 gccacactgctggaagaaaccccttaacaatcattgacattgagcagcttggcgagatt 1053
Db 1808 GCCACATCGCTGACGGAGAAATATGAGCGCTTAAGTGAAGAGAGTGTAGCGCTG 1867
Qy 1054 ctgagacaatttcattcgttctatgagcgccgataccaaagtagaanaagtgattctc 1111
Db 1868 CATATGCGCATCAAAATCTGACTTCTCTCAGATAGAACTAGAGTTGATGACCAATTTGTC 1927
Qy 1112 at-----cgaattatgactccaagagcaagaatgactgctgactgagactgag 1167
Db 1928 ATAAACCCACACATAAATTAACGGAAGAAACAAATGACATTGCGATGATCTTGA 1987
Qy 1168 aagcctctgacttcaacgacctagtgaaaccagtgctctgcccacaccagagcatgag 1227
Db 1988 ATGAAGTGAAGTACAGATTTATATACAGCTATTGTTTACGAGAAATGAATCAAGTT 2047
Qy 1228 ctgagacgaagacgctctgctgagcttcggggtgggggagccagagagaaggaag 1287
Db 2048 TTTCCCTCCAGAGAAATTTGTTCTATTTCTGCTGGGGGCGCACTTATATCAAGTTCT 2107
Qy 1288 acccagagagtgctggaagcgccagagtgctctcatttgagacaaagagcaagc 1347
Db 2108 ACTGAGAGCTGATGCAAGAGCTGACGTTCCCTTTCTATCAATGAGAAATGATCA---A 2164
Qy 1348 agatgtctctgacaacctgctcacacacagcaagctgctgctgctccctcagaggg 1407
Db 2165 CACAGATGCCAGATATACATTACGAGAAATATGCTGTGCGAGGCTATGACACAGGA 2224

Qy 1408 aacgtcattcttcgacaggtgagagtgagcgctctgctgacttcgaagaacatattc 1467
Db 2225 GGGTGAATTTCTTGTCAAGGGGAGTTCAAGCGGACACTGATGCTCAAGAAACACAGA 2284
Qy 1468 tgggtgctgataagagagatacaagctggtgtctgctgctgcaaaagcttacagacaga 1527
Db 2285 TGGCTCTGGCGGCGTGCATTTGGATATCATGTGCACCTGCGCAATGCGCCAGG 2344
Qy 1528 gtgtacgggaatgtgagtatttcacggactgcat 1562
Db 2345 GTGTATGCCCGGCTCCCAAGGTTCAACAGATGTGAT 2379

RESULT 4
US-08-508-448C-15
Sequence 15, Application US/08508448C
Patent No. 5804410
GENERAL INFORMATION:
APPLICANT: Kazuyoshi YAMAOKA et al.
TITLE OF INVENTION: NUCLEIC ACID SEQUENCE ENCODING
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.A.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/508,448C
FILING DATE: July 28, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX:
TELEX:
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1517 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: trachea
US-08-508-448C-15

Query Match 6.1%; Score 105.8; DB 2; Length 1517;
Best Local Similarity 49.9%; Pred. No. 1.4e-19;
Matches 356; Conservative 0; Mismatches 342; Indels 15; Gaps 3;
Qy 869 agagagagattgtgtggcgcgagagcgctcccggggctgtggcctggagagtgagcc 928
Db 612 AGCAGAGAACTCTTGGAGGCTGAGGCTGAGAGAGAGAGCTGCGCGCAAGTCAAGTC 671
Qy 929 tgcacgtccagagacgtgctgagaggtccatcatcaccccgagtgatcgtga 988

OY 415 YDLNLPAMCAAGFLOGNDSCGDSGGPLVTSKNMIM-WLIDGTSWGSCKAKAYRGVY 473

DB 668 TRLPFRDWI 677

OY 474 GNMVFTDWM 483

RESULT 8

ID 091674 PRELIMINARY; PRT: 1524 AA.

AC 091674;

DT 01-NOV-1996 (TEMBLrel. 01, Created)

DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)

DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)

DE POLYPROTEIN (OVOCHYMASE).

OS Xenopus laevis (African clawed frog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae; Xenopodidae; Xenopus.

RN (1)

RP SEQUENCE FROM N.A.

RL YANG J.C., LINDSAY L.L., HEDRICK J.L.;

RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE OF 1273-1524 FROM N.A.

RA YANG J.C., HEDRICK J.L.;

RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.

RL EMBL: U81290; AAC24717.1; -

DR EMBL: U44951; AAA91466.1; -

DR HSSP: P00763; IDPO.

DR PFAM: PF00431; CUB; 5.

DR PFAM: PF00089; trypsin; 3.

DR PRINTS: PR00722; CHYMOTRYPSIN.

KW Polypeptide.

FT CHAIN 57 308 SERINE PROTEASE.

FT CHAIN 584 817 SERINE PROTEASE.

FT CHAIN 1295 1524 OVOCHYMASE.

SO SEQUENCE 1524 AA; 167566 MW; 40D967CE CRC32;

- Query Match 18.4%; Score 677; DB 13; Length 1524;

Best Local Similarity 41.5%; Pred. No. 1,44e-147;

Matches 103; Conservative 54; Mismatches 79; Indels 12; Gaps 12;

DB 570 CGNAPTPKMLPRIVGEGEASPNMPOVIFLRTFCEGAIIISQWILTAHCT-RA 628

OY 243 CGVN-LNISR-QSRIVGESALPGAMPQVSLHVNHVCGSIIITPEWIVTAHCVEKP 300

DB 629 -APPSTVTVIAGD-NHR-MLNESTEQRNKTIRHNVNSEFYNDIALYLEEPDLN 685

OY 301 LNNPMHTAAGILROSFMFYGAGYOV-E-KVIS-HPNYSKTRKNDIALMLKQPLTFN 357

DB 686 DFNAPVCLPEPEVLTNPASVCVVTGNGNTAEDGQAPALGLOLPLIDSIICNTSY-YSG 744

OY 358 DLVAPVCLPRPGMLPQEOGLCSWGATEBEKRTSEVLNAAVLLEIQRCSRYIYDN 417

DB 745 ELTDHMLCAGPSSKEDACQDSGGGLVCQNEKEQFSYIGLVSWGEGCRVSPGYTK 804

OY 418 LITPMACAGFLQGNV-DSGQDSGGPLV-TSKNNIMWLGITSWSGCAKAYRPGVYGN 475

DB 805 VRLFTWI 812

OY 476 VMVFTDWM 483

RESULT 9

ID 088781 PRELIMINARY; PRT: 297 AA.

AC 088781;

DT 01-NOV-1998 (TEMBLrel. 08, Created)

DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)

DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)

DE SERINE PROTEASE PRECURSOR (FRAGMENT).

GN BSP2.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN-FISHER; TISSUE-BRAIN;

RX MEDLINE: 98389725.

RA DAVIES B.J., PICKARD B.S., STEEL M., MORRIS R.G., LATHE R.;

RT "Serine proteases in rodent hippocampus."

RL J. Biol. Chem. 273:23004-23011(1998).

DR EMBL: A2005642; CA06644.1; -

DR HSSP: P00763; IDPO.

DR PFAM: PF00089; trypsin; 1.

KW Signal: Protease; Serine protease.

FT NON_TER 1 1

FT SIGNAL <1 23 POTENTIAL.

FT CHAIN 24 297 SERINE PROTEASE.

SO SEQUENCE 297 AA; 32086 MW; 1B072619 CRC32;

- Query Match 17.0%; Score 625; DB 11; Length 297;

Best Local Similarity 36.3%; Pred. No. 3.42e-133;

Matches 87; Conservative 60; Mismatches 86; Indels 7; Gaps 5;

DB 38 NRYVGEDSDADQMPWIVSILKNGSHHAGSLTNRNVVSAACFSSNMKRPYSVLG 97

OY 253 SRIVGESALPGAMPQVSLHVNHVCGSIIITPEWIVTAHCVEKPLNNPMHTAFAG 312

DB 98 AMRLNGPSPQKRVGLASVLPFRYRKREKTHADILVRLERPIQSERILPLCPDSSV 157

OY 313 ILKQSEMF-YGAGYQVEKYSHPNYSK-TKNNDIALMLKQKFLTFENDLVKPCLEPDM 370

DB 158 HLPNNNCWTAAGSGTODGVLPRPOTLQKVPRIIDPELCSLYRAGAGEITEDMC 217

OY 371 MDPPEOLCWTISGWAGEEK-T-SEVLNAKAVLLETQCNRY--VYNDLITPAMIC 425

DB 218 AGYLEGRDACLDSGCPMLCQYDMLTLTGISWEGCAERNRPVYSLAHRPVOR 277

OY 426 AGFLQGNVDSGQDSGGPLVTSKNMIMWLGITSWSGCAKAYRPGVGNVFTDWMIR 485

RESULT 10

ID 043342 PRELIMINARY; PRT: 271 AA.

AC 043342;

DT 01-JUN-1998 (TEMBLrel. 06, Created)

DT 01-JUN-1998 (TEMBLrel. 06, Last sequence update)

DT 01-NOV-1999 (TEMBLrel. 12, Last annotation update)

DE SP0011A (FRAGMENT).

GN SP0011A.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

RN (1)

RP SEQUENCE FROM N.A.

RA RICKS D.O., BRUCE D., MUNDT M., DOGGERT N., MUNN C., SAUNDERS E.,

RA ROBINSON D., JONES M., BUCKINGHAM J., CHRISTEN L., THOMPSON S.,

RA GOODWIN L., BRYANT J., TESMER J., MEINKE L., LONGHIRE J., WHITE S.,

RA UENG S., TRITUM O., CAMPBELL C., FANCERT J., DEVERN L.;

RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA RICKS D.O., MAGNER R.P., MUNDT M.O.;

RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL: AC003965; AAB93671.1; -

DR HSSP: P00763; IDPO.

DR PFAM: PF00089; trypsin; 1.

FT NON_TER 1 1

SO SEQUENCE 271 AA; 29307 MW; FB404ADB CRC32;

- Query Match 16.6%; Score 608; DB 4; Length 271;

Best Local Similarity 37.4%; Pred. No. 1.64e-128;

Matches 89; Conservative 54; Mismatches 88; Indels 7; Gaps 5;

DB 2 NRYVGEDSDSDMPWIVSILKNGTHHAGSLTNRNVVTAACFSDNMKRPYSVLG 61

OY 253 SRIVGESALPGAMPQVSLHVNHVCGSIIITPEWIVTAHCVEKPLNNPMHTAFAG 312


```

OY 417 NLTPNACGFLGQNVDSGSGGGLVTSKNNIMWLIGDTSNGSCACARPRGYGNV 476
DB 321 SRFVTWIEGVNRNN 334
OY 477 MVFTDWIYRQMRAD 490

RESULT 14
ID 015146 PRELIMINARY; PRT; 810 AA.
AC 015146:
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DE 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE PLASMINOGEN PRECURSOR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN (1)
RP SEQUENCE FROM N.A.
RA BROWNE M.J., CHAPMAN C.G., DODD I., CAREY J.E., LAWRENCE G.M.,
RA MITCHELL D., ROBINSON J.H.;
RL Fibriolysis 0:0-0(0)
DR EMBL: M74220; AAA36451.1; -.
DR HSSP: P00747; 1PKR.
DR PFAM: PF00051; kringle; 5.
DR PFAM: PF00089; trypsin; 1.
DR PRINTS: PRO0018; KRINGLE.
DR PRINTS: PRO0722; CHYMOTRYPSIN.
KM Signal; zymogen.
FT SIGNAL 19 POTENTIAL.
FT CHAIN 20 810 PLASMINOGEN.
SQ SEQUENCE 810 AA; 90554 MW; 7E939182 CRC32;

Query Match 16.28; Score 595; DB 4; Length 810;
Best Local Similarity 39.48; Pred. No. 6,11e-125;
Matches 100; Conservative 53; Mismatches 89; Indels 12; Gaps 12;

DB 565 FDCGKPOVEBPKCPGRVYGCVAHPHSPMPQVSLRTFRGHPFGGTLISPEWVLAHCL 624
OY 241 IAGG-VNLNSR-QSRIVGESALPGAMPQVSLHV-NHYCGSGITTEPVIYTAHCV 297
DB 625 EKS-PRSSKYVILGA-HQEVNL-EPHQV-EIEVSLLEP-TRK-DIALKLSPAVIT 678
OY 298 EXPLNPMHMTARAGILRQSEMFYAGYQVEKYISHPNYDSKTKNDIALMKLQKPLTFN 357
DB 679 DAVIPACLPSPNVVADRTCEFTVGTGSETGTG-GAGLKEADLPVLEKVCN-RIEFLN 736
OY 358 DLVAPVCLPMPGMLOPEQLCMWISGATEKEKGTSEVLNAKVLLETORCNSRYVY-D 416
DB 737 GRVOSTELCAGHLAGSDSGSGGLVCFEEMKXYILOGVTSNGLGCARPNKPGVYVRY 796
OY 417 NLTPNACGFLGQNVDSGSGGGLVTSKNNIMWLIGDTSNGSCACARPRGYGNV 476
DB 797 SRFVTWIEGVNRNN 810
OY 477 MVFTDWIYRQMRAD 490

RESULT 15
ID 09XSM2 PRELIMINARY; PRT; 273 AA.
AC 09XSM2:
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DE 01-NOV-1999 (TREMblrel. 12, Last annotation update)
DE TRYPTASE (EC 3.4.21.55).
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
OC Caprinae; Ovis.
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE-ABOMASUM;

```

```

RA PEMBERTON A.D., MCALBESSE S.M., HUNTLEY J.F., MACKELLAR A.,
RA COLLIE D.D.S., MCWILLAN L., SCODAMORE C.L., MILLER H.R.P.;
RT "DNA sequence of sheep mast cell tryptase and its immunocalisation
RT in lung, skin and gut in comparison with sheep mast cell proteinase-
RT 1."
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: Y18224; CAB41989.1; -.
KM Hydrolyase.
FT CHAIN 29 TRYPTASE.
SQ SEQUENCE 273 AA; 30288 MW; 268A4B33 CRC32;

Query Match 15.98; Score 585; DB 6; Length 273;
Best Local Similarity 39.58; Pred. No. 3.38e-122;
Matches 98; Conservative 48; Mismatches 84; Indels 18; Gaps 14;

DB 25 QRSGITGKREA-PGSRMPQVSLRVDRQYWRHOCGSLHPQVLAHACIGPELOEP-- 81
OY 251 KQSRIVGESALPGA-WPMQVSLHV-QN-VHYCGSGITTEPVIYTAHCVKPLNPNH 306
DB 82 -SDFRVQLREQHLHYODRLPLISRVIPHPHYVVENGADIALQLEBPVSRHQVPTL 140
OY 307 WTAFAGILRQSEMFYAGY-QVEKYISHPNYDSKTKNDIALMKLQKPLTFNDLVKPVCL 365
DB 141 PPASETFPPESSQCVWVGKDVNGRPLPPYPLKQYKVPVENSVCDDKY-HSGLETDIS 199
OY 366 PNPMMLOPEQLCMWISGATEE-KGTSE-VLNAKVLLETORCNSRYVDNLITP-A 422
DB 200 VPIVQEDNLACGDSGRDSCGDSGGLVCFEEMKXYILOGVTSNGLGCARPNRGIYTRIT 259
OY 423 M-IC-AGFL-QGNV--DSCGDSGGLVTSKNNIMWLIGDTSNGSCACARPRGYGNV 477
DB 260 SYLDWIRQ 267
OY 478 VFTDWIYR 485

```

Search completed: Thu Apr 20 00:00:33 2000
Job time : 914 secs.

OY 1216 ccaggcatgctgcagcccaagacagctctgattccggtgagggccaccag 1275
 DB 1020 tctgagagagacgctgctctgctgcttctcattgctgagggccagctgctg 1079
 OY 1276 gagaagagggaagacctcagaagtctgaaagctgcgaagtcttcattcattagacag 1335
 DB 1080 gacctgtggccagccctgagctgagctgagctgacgctgacgctgagctgagctgag 1139
 OY 1336 agatgc-----aacagcagatatgtctatgacaaactgacacccacatgac 1386
 DB 1140 gacctgctgcacagctgacagagagtgagagctgctcccaaatatcagagatgattc 1199
 OY 1387 tctgctgctgctcctgcagggggaagctgattctctgcagagtgagagtgagctctg 1446
 DB 1200 tctgctgctgctgctgagctgagctgagctgagctgagctgagctgagctgagctg 1259
 OY 1447 gtcactgcgaagaaacatactgctgctgctgctgctgctgctgctgctgctgctg 1506
 DB 1260 gccaccactgacacggggcagctgctgctgctgctgctgctgctgctgctgctgctg 1319
 OY 1507 gccaaagcttacagaccagagtgctgagggagtgatgctgctgctgctgctgctgctg 1566
 DB 1320 gccaccgtggggcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1379
 OY 1567 cgacaaatgagggcaga 1583
 DB 1380 aagctcattgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1396

RESULT 9

US-08-321-777-3
 ; Sequence 3, Application US/08321777
 ; Patent No. 5504067
 ; GENERAL INFORMATION:
 ; APPLICANT: MORRISSEY, James H.
 ; APPLICANT: Comp. Philip C.
 ; TITLE OF INVENTION: Treatment of Bleeding with Modified
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESSES:
 ; ADDRESSEE: Richards, Medlock & Andrews
 ; STREET: 1201 Elm Street, Suite 4500
 ; CITY: Dallas
 ; STATE: Texas
 ; COUNTRY: US
 ; ZIP: 75270-2197
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/321.777
 ; FILING DATE:
 ; CLASSIFICATION: 514
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/882202
 ; FILING DATE: 13-MAY-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Hansen, Eugenia S.
 ; REGISTRATION NUMBER: 31,966
 ; TELEPHONE/DOCKET NUMBER: OMRF B34290C
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 214-939-4500
 ; TELEFAX: 214-939-4600
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1440 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA

HYPOTHETICAL: NO
 AMTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 TISSUE TYPE: Blood
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 36..1433
 OTHER INFORMATION:
 OTHER INFORMATION: factor VII cDNA
 US-08-321-777-3

Query Match 5.5%; Score 95.4; DB 1; Length 1440;
 Best Local Similarity 47.9%; Pred. No. 9.4e-17;
 Matches 353; Conservative 0; Mismatches 366; Indels 18; Gaps 2;

OY 865 ccagcagagcagagatgtggtggcgagagcgctcccgaggcgctgctgctgctgctgctg 924
 DB 660 ccccaagcgcgaattgctggggcgaagctgctccccaagggagtgctcattgctgctgctg 719
 OY 925 agccctgacgctcagagagctcagcgtgtgctgagagctcattcattcaccacccgagtgatc 984
 DB 720 cgtgtgttggtgaatgacactgattgtgtgggggagcctgattcaacacactgctgctg 779
 OY 985 gtgacagcgcgcactgtgtggaagaaacctcttaacatccatgcatgagcagcatt 1044
 DB 780 gtctccgcccacactgttctgctgctgctgctgctgctgctgctgctgctgctgctgctg 839
 OY 1045 gcgggagatttgagaaatcttcaatgctgctgctgctgctgctgctgctgctgctgctgctg 1104
 DB 840 ggagagcagacactgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgac 899
 OY 1105 attctcaccacaaatatagtactccaaagacaaagaaatgacattgtgctgattgagctg 1164
 DB 900 atcattcccccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 959
 OY 1165 cgaagacgtcctgacttcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1215
 DB 960 caccacccgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1019
 OY 1216 ccaggcatgctgcagcccaagacagctctgattccggtgagggccaccagag 1275
 DB 1020 tctgagagagacgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1079
 OY 1276 gagaagagggaagacctcagaagtctgaaagctgcgaagtcttcattcattagacag 1335
 DB 1080 gacctgtggccagccctgagctgagctgagctgagctgagctgagctgagctgagctgag 1139
 OY 1336 agatgc-----aacagcagatatgtctatgacaaactgacacccacatgac 1386
 DB 1140 gacctgctgcacagctgacagagagtgagagctgctcccaaatatcagagatgattc 1199
 OY 1387 tctgctgctgctcctgcagggggaagctgattctctgcagagtgagagtgagagtgagctctg 1446
 DB 1200 tctgctgctgctgctgagctgagctgagctgagctgagctgagctgagctgagctgag 1259
 OY 1447 gtcactgcgaagaaacatactgctgctgctgctgctgctgctgctgctgctgctgctg 1506
 DB 1260 gccaccactgacacggggcagctgctgctgctgctgctgctgctgctgctgctgctg 1319
 OY 1507 gccaaagcttacagaccagagtgctgagggagtgatgctgctgctgctgctgctgctg 1566
 DB 1320 gccaccgtggggcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1379
 OY 1567 cgacaaatgagggcaga 1583
 DB 1380 aagctcattgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1396

RESULT 10

PCF-US93-04493-3
 ; Sequence 3, Application PC/TUS9304493

```

Db 279 AALALVDKVCVCTGTMGNTQFGQAAVLDQARVPIISNEVCNSPDPYQNIKPKKECA 338
OY 367 NPGMTLQPEQLCWISSMGATEEKGTSEVLAANKVLLIETQNCNSRYIYDNLTPAMICA 426
Db 339 GYEGGCIADCGSGGSHFYCEDRISTGTSNRLCGIVSMGTGCALARKPGVYTKYIDREW 398
OY 427 GFLGDNDSCQSGSGGRLV-TSK-NNI--WMLIGDTSMNSGCAKAYRRGVYGNMVFETDW 482
Db 399 IFQAIKTHS 407
OY 483 IYROMRADG 491

RESULT 11
ENTRY PLMS #type complete
TITLE plasmin (EC 3.4.21.7) precursor - mouse
CONTAINS angiotatin; plasminogen
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 20-Sep-1991 #sequence_revision 01-Nov-1996 #text_change 18-Jun-1999

ACCESSIONS A38514; S48202; S48203
REFERENCE A38514
#authors Deegen, S.J.F.; Bell, S.M.; Schaefer, L.A.; Elliott, R.W.
#journal Genomics (1990) 8:49-61
#title Characterization of the cDNA coding for mouse plasminogen and
#molecule_type mrna localization of the gene to mouse chromosome 17.
#accession A38514
#residues 1-812 #label DEG
#cross-references GB:004766; NID:g200402; PIDN:AAA50168.1; PID:g200403
S48202
#authors Lijnen, H.R.; van Hoef, B.; Beelen, V.; Collen, D.
#journal Eur. J. Biochem. (1994) 224:863-871
#title Characterization of the murine plasma fibrinolytic system.
#cross-references NUID:95010076
#accession S48202
#molecule_type protein
#residues 20-25 #label LIT
#accession S48203
#molecule_type protein
#residues 22-27 #label LIT
COMMENT Plasminogen is synthesized by the kidney and is present in plasma
and many other extracellular fluids.
COMMENT Plasminogen is converted into plasmin by plasminogen activators,
both plasminogen and its activator being bound to fibrin. Plasmin
is inactivated by alpha-2-antiplasmin (see PIR:S47217)
immediately after dissociation from the clot. In the presence of
the inhibitor, the activation involves only cleavage after
Arg-581, resulting in two chains connected by two disulfide
bonds. Without the inhibitor, the activation involves also
removal of the activation peptide.
COMMENT Streptolysin 1 (see PIR:KMSS1) acts on plasminogen to produce
angiotatin. Together with endostatin (see PIR:A56101,
PIR:B56101), angiotatin acts to inhibit angiogenesis, and so may
be useful in treating solid tumors.

FUNCTION
#description dissolves the fibrin of blood clots; acts as a proteolytic
factor in a variety of processes including embryonic
development, tissue remodeling and tumor invasion; in
ovulation it weakens the walls of the graafian follicle;
also activates the urokinase-type plasminogen activator
fibrinolysis
#superfamily plasmin; kringie homology; plasminogen-related
protein precursor homology; trypsin homology
#protein_inhibitor blood; duplication; fibrinolysis;
glycoprotein; hydrolase; kidney; kringie; plasma; serine
proteinase; zymogen

CLASSIFICATION
#pathway
KEYWORDS
FEATURE
1-96 #domain plasminogen-related protein precursor homology
1-19 #label PLPV
1-19 #domain signal sequence #status predicted #label SIG
20-812 #product plasminogen #status predicted #label PRO

```

Accessions	Author	Title	Year	Accessions	Author	Title	Year
20-66		domain activation peptide	1986	20-66		domain activation peptide	1986
79-666		product angiotensin	1986	79-666		product angiotensin	1986
97-581, 582-812		product plasmin	1986	97-581, 582-812		product plasmin	1986
97-581		domain chain A	1986	97-581		domain chain A	1986
103-181		domain kringles	1986	103-181		domain kringles	1986
185-262		domain kringles	1986	185-262		domain kringles	1986
275-352		domain kringles	1986	275-352		domain kringles	1986
377-454		domain kringles	1986	377-454		domain kringles	1986
481-560		domain kringles	1986	481-560		domain kringles	1986
582-812		domain chain B	1986	582-812		domain chain B	1986
582-805		domain trypsin	1986	582-805		domain trypsin	1986
49-73, 53-61,		103-181, 124-154,		49-73, 53-61,		103-181, 124-154,	
152-176, 185-252,		188-316, 206-245,		152-176, 185-252,		188-316, 206-245,	
234-257, 275-352,		296-335, 324-347,		234-257, 275-352,		296-335, 324-347,	
377-454, 398-437,		426-449, 481-560,		377-454, 398-437,		426-449, 481-560,	
502-543, 553-555,		568-687, 578-586,		502-543, 553-555,		568-687, 578-586,	
609-625, 701-768,		731-747, 758-786		609-625, 701-768,		731-747, 758-786	
78-79				78-79			
136, 308				136, 308			
466-467				466-467			
581-582				581-582			
624, 667, 762				624, 667, 762			
SUMMARY				SUMMARY			
Query Match				Query Match			
Best Local Similarity 41.4%				Best Local Similarity 41.4%			
Matches 111; Conservative 51; Mismatches 93; Indels 13; Gaps 12;				Matches 111; Conservative 51; Mismatches 93; Indels 13; Gaps 12;			
Db 553 DYCDIPICASASFECKGPOVERPKCGRGVAGCVANPHSPMPQISLRTRFGQHFCCGT 612				Db 553 DYCDIPICASASFECKGPOVERPKCGRGVAGCVANPHSPMPQISLRTRFGQHFCCGT 612			
Qy 228 DACSKAVVSLRCLACG-VNLNSR-QSRIVGESALLPGAMPQVSLHVG-NV-HVCGGS 283				Qy 228 DACSKAVVSLRCLACG-VNLNSR-QSRIVGESALLPGAMPQVSLHVG-NV-HVCGGS 283			
Db 613 LIAPENVLTAAHCLERS-SRP-EP--YKVL-CAHEEYIRGLDVGQ-ISAALLILEPNRR 666				Db 613 LIAPENVLTAAHCLERS-SRP-EP--YKVL-CAHEEYIRGLDVGQ-ISAALLILEPNRR 666			
Qy 284 IITPEWIVTAAHCEKELNPNMHWTAFAGLRQSFMYGAGYQVERVYISHPNYDSKTKNN 343				Qy 284 IITPEWIVTAAHCEKELNPNMHWTAFAGLRQSFMYGAGYQVERVYISHPNYDSKTKNN 343			
Db 667 DIALTKSRPTINDKYIPACLPSPNPMVADRITCYITGGETGQFFGAGR-LKEADLPV 725				Db 667 DIALTKSRPTINDKYIPACLPSPNPMVADRITCYITGGETGQFFGAGR-LKEADLPV 725			
Qy 344 DIALTKSRPTINDKYIPACLPSPNPMVADRITCYITGGETGQFFGAGR-LKEADLPV 725				Qy 344 DIALTKSRPTINDKYIPACLPSPNPMVADRITCYITGGETGQFFGAGR-LKEADLPV 725			
Db 726 IENRYCN-RVEYLNNRYKSTELACAGLAGVDSQCGDSGGPLVCEKDYKYLIGQVTSNGL 784				Db 726 IENRYCN-RVEYLNNRYKSTELACAGLAGVDSQCGDSGGPLVCEKDYKYLIGQVTSNGL 784			
Qy 404 IENRYCN-RVEYLNNRYKSTELACAGLAGVDSQCGDSGGPLVCEKDYKYLIGQVTSNGL 784				Qy 404 IENRYCN-RVEYLNNRYKSTELACAGLAGVDSQCGDSGGPLVCEKDYKYLIGQVTSNGL 784			
Db 785 GCAPNPKPGYTVRVSRRVPIERMRNN 812				Db 785 GCAPNPKPGYTVRVSRRVPIERMRNN 812			
Qy 463 GCAPNPKPGYTVRVSRRVPIERMRNN 812				Qy 463 GCAPNPKPGYTVRVSRRVPIERMRNN 812			
RESULT 12				RESULT 12			
ENTRY				ENTRY			
TITLE				TITLE			
ALTERNAME_NAMES				ALTERNAME_NAMES			
CONTAINS				CONTAINS			
ORGANISM				ORGANISM			
DATE				DATE			
ACCESSIONS				ACCESSIONS			
REFERENCE				REFERENCE			
authors				authors			
#journal							

```

APPLICATION NUMBER: 08/065,725
FILING DATE: 21-MAY-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/662,920
FILING DATE: 28-FEB-1991
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Parmelee, Steven W.
REGISTRATION NUMBER: 31,990
REFERENCE/DOCKET NUMBER: 13952-8-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-467-9600
TELEFAX: 415-543-5043
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2422 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: N
ANTI-SENSE: N
FEATURE:
NAME/KEY: CDS
LOCATION: 28..1420
OTHER INFORMATION: /codon_start= 28
OTHER INFORMATION: /product="Factor VII"
US-08-475-845-1

```

```

Query Match          5.5%; Score 95.4; DB 2; Length 2422;
Best Local Similarity 47.9%; Pred. No. 1,2e-16;
Matches 353; Conservative 0; Mismatches 366; Indels 18; Gaps 2;

```

```

QY 865 cgcacgaagcagaattgctggcgcgagagcgctcccggggcccctggccctggcagctc 924
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 539 CCCCAAGCGCAATTGTGGGGGCAAGGTGTGCCCCCAAGGGAGGTGCTCATGTGGCAGTGC 658
QY 925 agccgcagctcagacagctccacggtgctcgagagctccatcatcccccggcggtgac 944
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 659 CTGTTGTGTGAATGTGAGCACTGTTGTGGGGGAGCCCTGATCAACACCATCTGGGTG 718
QY 985 gtacagacccgcacactgctggaagaacctcttaacatccatcagatcgagcagatt 1044
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 719 GTCTCCGGGGCCCACTGTTTCCAGCAAAATCAGAACTGAGAGAACTGATCGCGGTGCTG 778
QY 1045 gcggggattttgagacaattcttcatgttctatgtgagcgagataccaagtagaagaa 1104
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 779 GCGGAGCAGACCTGACGAGACGACGAGGGGATGACAGACGCGCGGGGTGCGCAGGTC 838
QY 1105 attctcatcaaatatgtactccaagaaccaagaacaatgacttgcgtgtgagagctg 1164
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 839 ATCATCTCCAGCAGCTACCTCCCGGGCACACACACAGCACTCCGCTGCTCGCCCTG 898
QY 1165 cggaaagcctctgacttcaagcagctagtagaaaccagtggtgtgtgcccac-----c 1215
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 899 CACCAACCCGCTGCTCACTGACATGTGTGTCCTCTGCTGCTGCTGCTGCTGCTGCTG 958
QY 1216 ccaagcatatgtctcagcagcagaacagctctcgtgatttcccggtcggggggcgag 1275
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 959 TCTGAGAGAGACGCTGCGCTTCTGTCGCTTCTCATTTGTGTGAGCGGCTGGGCCAGTGTG 1018
QY 1276 gagaagaaggaagcctcagaagtgtcgaagcgtcccaagtgcttctcatgtgagaacg 1335
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1019 GACCGTGGCGCCACGGCCCTGTGAGCTCATGTCTCTACGTGCGCCCGCGCTGATGCCAG 1078
QY 1336 agatgc-----aacagcagatatgtctatgacaacctgataccacagcagatc 1386
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1079 GACTGCTCTCAGCAGTCAAGGAGGTGGAGACTCCCCCAAAATATCAGGGAGTCAATGTTTC 1138
QY 1387 tctgcaggtctcctcaggggaagcgtcgtattcttgcaggggtgacagtggaagggcctctg 1446
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

DB 1139 TGTGCGGCTACTCGATGCGACAGCAAGACTCTCTCAAGGGGAGACTGAGGCCACAT 1198
QY 1447 gtacttcgaagaacaatatctgtgtgctgataaggagatacaagctgggtgtcgtgtc 1506
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1199 GCCACCACTACCGGGGACGCGGTACTGACAGGGGCACTGCTAGCTGGGGCCAGGGCTGC 1258
QY 1507 gccaaagcttacagaccagagagtgacgggaatgtgagtgttctaacggagctgattat 1566
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1259 GCACCGTGGGCGCACTTGTGGGTGTACACAGGCTCTCCCACTACATGAGTGGCTGCA 1318
QY 1567 cgacaatgagggcaga 1583
DB 1319 AAGCTCATGCGCTCAGA 1335

```

```

RESULT 12
US-08-327-690-1
Sequence 1, Application US/08327690
Patent No. 5817788
GENERAL INFORMATION:
APPLICANT: Berkner, Kathleen L.
APPLICANT: Petersen, Lars C.
APPLICANT: Hart, Charles E.
APPLICANT: Hedner, Ulla
APPLICANT: Bregengaard, Claus
TITLE OF INVENTION: Modified Factor VII
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
STREET: Townsend and Townsend Kourile and Crew
STREET: One Market Plaza, Steuart Street Tower
CITY: San Francisco
STATE: CA
COUNTRY: U.S.A.
ZIP: 94105-1492
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/327,690
FILING DATE: 24-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/065,725
FILING DATE: 21-MAY-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/662,920
FILING DATE: 28-FEB-1991
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Parmelee, Steven W.
REGISTRATION NUMBER: 31,990
REFERENCE/DOCKET NUMBER: 13952-8-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-467-9600
TELEFAX: 415-543-5043
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2422 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: N
ANTI-SENSE: N
FEATURE:
NAME/KEY: CDS
LOCATION: 28..1420
OTHER INFORMATION: /codon_start= 28
OTHER INFORMATION: /product="Factor VII"
US-08-327-690-1

```

```

Db 839 ATCATCCCAACAGCTACGTCCTCCGGGACACCAACCAACGACATCGCGCTCTCCGCTG 898
Oy 1165 cagaagcctctgacttcaagaagctagtgaaaccagtggtctgcccac-----c 1215
Db 899 CACCAAGCCCGTGTCTCTACTGACCATGTGTGCCCCCTCTGCTGCCCCGAACGACGTTTC 958
Oy 1216 ccaagcctgactgctgcaagcacaagcctctgctgacttccgggtgaggggccaccag 1275
Db 959 TCTGAGAGGAGCGTGGGCTGTGCTGCTTCTCATTTGGTCACAGGCTGGGCTGAGCTGCTG 1018
Oy 1276 gagaagagggaagacctcaagaagtctgaagcctcacaagtgcttctcatggagaacag 1335
Db 1019 GACCGTGGCGGCGGCGGCTGTGAGCTCATGTCTCCTCAACGTCGCCCGGCTATGACCCAG 1078
Oy 1336 agatgc-----aaacagagatatgtctatgacaaaccgcatcacacacagctgac 1386
Db 1079 GACTGCTGCACAGTACAGGACGAGGTGGAGAGCTCCCAAAATATCAGCACTACTGTTTC 1138
Oy 1387 tgtgcgagcttctcgcagagggaagctgacttctgcaagggtgacagtgaggggctctg 1446
Db 1139 TGTGCTGGCTACTGTGGATGGCAGCAGACTCTGCAAGGGGGACAGTGGAGGCCACAT 1198
Oy 1447 gtcacttcgaagaacaatatctgtgtgctgtagggaggtacaaagctgggtctgctgt 1506
Db 1199 GCCACCCACTACCGGGGACGCTGTACTGTGACGGGCACTGTACAGTGGGGCCAGGGCTGC 1258
Oy 1507 gccaaagcttcaagaccagaggtgtacgggaagtgtgatatcaccagactgtattat 1566
Db 1259 GCAACCCGTGGGCGCACTTTGGGGGTGACACCAAGGCTCTCCAGTACATGAGTGGCTGCA 1318
Oy 1567 cgacaatgagggcaga 1583
Db 1319 AAGCTCATGCGCTCAGA 1335

```

RESULT 14

```

US-08-537-807-1
; Sequence 1, Application US/08537807
; Patent No. 5861374
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Modified Factor VII
; NUMBER OF SEQUENCES: 4
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/537,807
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/05779
; FILING DATE: 23-MAY-1994
; APPLICATION NUMBER: US 08/065,725
; FILING DATE: 21-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/662,920
; FILING DATE: 28-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Parmelee, Steven W.
; REGISTRATION NUMBER: 31,990
; REFERENCE/DOCKET NUMBER: 13952-8-1PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-467-9600
; TELEFAX: 415-543-5043
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2422 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: cDNA
; HYPOTHETICAL: N
; ANTI-SENSE: N
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 28..1420
; OTHER INFORMATION: /codon_start= 28
; OTHER INFORMATION: /product= "Factor VII"
US-08-537-807-1

```

```

Query Match 5.58; Score 95.4; DB 3; Length 2422;
Best Local Similarity 47.9%; Pred. No. 1,2e-16;
Matches 353; Conservative 0; Mismatches 366; Indels 18; Gaps 2;

```

```

Oy 865 cgcacagagcagattgtggtgagcgagagcgcgctccgggggctgagcctgagagtg 924
Db 599 CCCCAGGCCCAATTGTGGGGCAAGGTGTGCCCCAAAGGGAGTGTCTCATGGCGAGGTC 658
Oy 925 agcctgcagctccagaaagctcacaagtggtgcgagagctccatcatcaccccgagtgatc 984
Db 659 CTGTTGTGTGTAATGAGCTCAAGTGTGTGGGGGAGACCTGATCAACCATCTGGGTG 718
Oy 985 gtgacagcgccgcacatcggtgtgaaaaaacctcttaacaaaccatgagcattggaacg 1044
Db 719 GTCGCCGGCGGCCCACTGTTTTCGACAAATCAAGAACTGAGAGAACTGATCGCGGTGCTG 778
Oy 1045 gctggagatttgaagacaaatcttcatgttcatgtgagccgagataccagtgaaaaagt 1104
Db 779 GCGGAGCAGCACTCAGCGAGCAGACGAGGAGTGAAGCAGACCGCGGGTGGCCAGGTC 838
Oy 1105 attctcatccaattatgacttccaaagcacaagaacaatgacatgtgcgtatgaagctg 1164
Db 839 ATCATCCCAACAGCTACGTCCTCCGGGACACCAACCAACGACATCGCCTCTCCGCTG 898
Oy 1165 cagaagcctctgacttcaagaagctagtgaaaccagtggtctgcccac-----c 1215
Db 899 CACCAAGCCCGTGTCTCTACTGACCATGTGTGCCCCCTCTGCTGCCCCGAACGACGTTTC 958
Oy 1216 ccaagcctgactgctgcaagcacaagcctctgctgacttccgggtgaggggccaccag 1275
Db 959 TCTGAGAGGAGCGTGGGCTGTGCTGCTTCTCATTTGGTCACAGGCTGGGCTGAGCTGCTG 1018
Oy 1276 gagaagagggaagacctcagaagtgctgaagcctgcaagtgcttctcatggagaacag 1335
Db 1019 GACCGTGGCGGCGGCGGCTGTGAGCTCATGTCTCTCAACGTCGCCCGGCTATGACCCAG 1078
Oy 1336 agatgc-----aaacagagatatgtctatgacaaaccgcatcacacacagctgac 1386
Db 1079 GACTGCTGCACAGTACAGGACGAGGTGGAGAGCTCCCAAAATATCAGCACTACTGTTTC 1138
Oy 1387 tgtgcgagcttctcgcagagggaagctgacttctgcaagggtgacagtgaggggctctg 1446
Db 1139 TGTGCTGGCTACTGTGGATGGCAGCAGACTCTGCAAGGGGGACAGTGGAGGCCACAT 1198
Oy 1447 gtcacttcgaagaacaatatctgtgtgctgtagggaggtacaaagctgggtctgagcgt 1506
Db 1199 GCAACCCACTACCGGGGACGCTGTACTGTGACGGGCACTGTACCTGGGCGGCGAGGCTGC 1258
Oy 1507 gccaaagcttcaagaccagaggtgtacgggaagtgtgatatcaccagactgtattat 1566
Db 1259 GCAACCCGTGGGCGCACTTTGGGGGTGACACCAAGGCTCTCCAGTACATGAGTGGCTGCA 1318
Oy 1567 cgacaatgagggcaga 1583
Db 1319 AAGCTCATGCGCTCAGA 1335

```

RESULT 15

```

US-08-871-003-1
; Sequence 1, Application US/08871003
; Patent No. 5997864
; GENERAL INFORMATION:

```


membrane attachment using a signal-anchor sequence.
conversion from membrane-bound to soluble forms may involve further processing. See the human entry (PIR:A56318) for an alternative version of products.
mature enteropeptidase is variously reported to contain two (heavy and light) or three (mini, heavy, and light) chains derived from a single precursor form; heavy and light chains are disulfide linked

COMPLEX

FUNCTION
#description
#pathway
#classification

cleaves propeptide from trypsinogen to produce active trypsin
intestinal digestive hydrolase cascade
#superfamily enteropeptidase; Clr/Cis repeat homology; LDL receptor ligand-binding repeat homology; MAM homology; scavenger receptor cysteine-rich domain homology; trypsin homology
glycoprotein; hydrolase; intestine; serine proteinase; transmembrane protein; zymogen

KEYWORDS

FEATURE
22-38
52-117
118-800
199-236
358-530
542-647
653-653
694-799
801-1035

#domain transmembrane #status predicted #label TMN\
#product enteropeptidase mini chain #status predicted
#label MCH\
#product enteropeptidase heavy chain #status predicted
#label HCH\
#domain LDL receptor ligand-binding repeat homology
#label LDL1\
#domain MAM homology #label MAM\
#domain Clr/Cis repeat homology #label Clr\
#domain LDL receptor ligand-binding repeat homology
#label LDL2\
#domain scavenger receptor cysteine-rich domain homology
#status atypical #label SRCR\
#product enteropeptidase light chain #status predicted
#label LCH\
#domain trypsin homology #label TRY\
801-1030
116,147,170,194,
233,263,264,404,
456,486,519,550,
646,698,722,741,
762,864,903,965
#binding-site carbohydrate (Asn) (covalent) #status predicted\
788-912,826-942,
926-993,957-972,
983-1011
841,892,987
#length 1035 #molecular-weight 114866 #checksum 4578
SUMMARY

Query Match 24.4%; Score 895; DB 1; Length 1035;
Best Local Similarity 36.3%; Pred. No. 3,16e-180;
Matches 138; Conservative 84; Mismatches 141; Indels 17; Gaps 16;

DB 659 CKEDNFOCKD-GECIPLVNLCDFPHCKGSDAHCVRLFNGTTDSSGLVQFRIQSI-WH 716
113 GNSNGECSSGCTCINPSMNCDSVSHCPGEGEDNRCVRLT-GP-NFI-LOVYSQKRSWH 169
DB 717 VACAEHWMTQISDVCOLIGTG-NSSVPTFS-TGGCPVYNLTFAP-NSLLILT-PSQO 772
170 PWCODHNEWGAACRDKYKNNFYSSOGIVDSGSTSMKNTSGANVDIYKKHYSDA 229
DB 773 CLEDSILLQCNKSCGKLVTOEVSPIKIVGSDSREGANPVVALYFDDQVCGASLVS 832
230 GSKKAVVSLRCI--ACGVNINSSROS-RIVGESALPGAMPQVSLHVQNVHVGCGSIIT 286
DB 833 KDLVLSAARHYGRNNE-PSKRAVILGLHMASNLTPQIETRLIDQIVINPHYKRRKN 891
287 PEMIVYPAARCV-EKPLNPNHMTAFAGILKROSFMEYGA-GYQ-VEKVISPNYDSKTKNN 343
DB 892 DIAMHLEKVVNTDYLOPCLPEENOVFPFGRTCSLAGEALYOGSTADVLQADAVPL 951
344 DIALMKQKLTNDLVKPCLPKPMLOPBDLCWISGNGATPEKKTSEVLNAKAVLL 403
DB 952 LSNKCOQO-MPEYNITENNVCAGYENGVDSCGDSGGLPKCEGNRMILLACTSTGYQ 1010
404 IETORCSRYVDNLITPAMICAGFLQGNVDSGCGSLVTSKNNIMWLLIGDTSWGSG 463

DB 1011 CALPNRPVYARVPRFTENI 1030
464 CAKAYRPGYGVNMTETDWT 483

RESULT 2
ENTRY
TITLE
ALTERNATE_NAMES
ORGANISM
DATE

A56318 #type complete
enteropeptidase (EC 3.4.21.9) precursor - human
enterokinase
#formal_name Homo sapiens #common_name man
19-May-1995 #sequence_revision 09-Aug-1996 #text_change
18-Jun-1999
A56318; B43090

ACCESSIONS
REFERENCE
#authors
#journal
#title

Kitamoto, Y.; Velle, R.A.; Donis-Keller, H.; Sadler, J.E.
Biochemistry (1995) 34:4562-4568
cDNA sequence and chromosomal localization of human
enterokinase, the proteolytic activator of trypsinogen.
#accession A56318
#molecule_type mRNA
#residues 1-1019 #label KIT
#cross-references GB:U09860; NID:G746412; PIDN:AAC50138.1; PID:G746413
REFERENCE
#authors
#journal
#title

A43090
Kitamoto, Y.; Yuan, X.; Wu, Q.; McCourt, D.W.; Sadler, J.E.
Proc. Natl. Acad. Sci. U.S.A. (1994) 91:7588-7592
Enterokinase, the initiator of intestinal digestion, is a
mosaic protease composed of a distinctive assortment of
domains.
#cross-references NUID:94329561
#accession B43090
#molecule_type mRNA
#residues 749-1019 #label KIT2
#cross-references GB:U09860
#residues 749-1019 #label KIT2
COMMENT
The mechanism of association with the membrane of the intestinal
brush border is unknown. The sequence is compatible with type II
membrane attachment using a signal-anchor sequence (as annotated
below) or with amino-terminal myristoylation of the heavy chain.

GENETICS
#gene
#cross-references GDB:384083; OMIM:226200
#map_position 21q21-21q21
COMPLEX
#description
#pathway
#classification

GDB:PRS57
#cross-references GDB:384083; OMIM:226200
#map_position 21q21-21q21
Mature enteropeptidase is variously reported to contain two (heavy and light) or three (mini, heavy, and light) chains derived from a single precursor form; heavy and light chains are linked by a disulfide bond. Possibly, conversion from membrane-bound to soluble forms involves further processing. See the pig or bovine entries (PIR:A53663, PIR:A43090) for an alternative version of products.

FUNCTION
#description
#pathway
#classification

cleaves activation peptide from trypsinogen to produce active trypsin
intestinal digestive hydrolase cascade
#superfamily enteropeptidase; Clr/Cis repeat homology; LDL receptor ligand-binding repeat homology; MAM homology; scavenger receptor cysteine-rich domain homology; trypsin homology
glycoprotein; hydrolase; intestine; serine proteinase; transmembrane protein; zymogen

KEYWORDS

FEATURE
1-784
22-38
184-221
342-504
526-631
643-677
678-783

#product enteropeptidase heavy chain #status predicted
#label HCH\
#domain transmembrane #status predicted #label TMN\
#domain LDL receptor ligand-binding repeat homology
#label LDL1\
#domain MAM homology #label MAM\
#domain Clr/Cis repeat homology #label Clr\
#domain LDL receptor ligand-binding repeat homology
#label LDL2\
#domain scavenger receptor cysteine-rich domain homology
#status atypical #label SRCR\
#status atypical #label SRCR

QY 773 tctataaaaaactgtc 787
 DB 15 TTTATATAAAACTGT 1

RESULT 5
 T79127
 ID T79127 standard; cDNA to mRNA; 699 BP.
 AC T79127;
 DT 08-OCT-1997 (first entry)
 DE Human serine protease 60 (SP60) cDNA.
 KM Human; colon carcinoma; COLO 201; cell line; serine protease; SP60;
 OS screening; inhibitor; treatment; disease; ds.
 FH Homo sapiens.
 FT Key Location/Qualifiers
 FT mat_peptide 1..699
 FT /tag=a
 FT /product=serine_protease
 FT J09149790-A.
 PN 10-JUN-1997.
 PD 24-JUL-1996; 212196.
 PE 29-SEP-1995; JP-275105.
 PR (SUNR) SUNTORY LTD.
 PA WPI; 97-357902/33.
 DR P-PSDB; W22986.
 PT Human colon carcinoma derived serine protease(s) SP59, SP60 and SP67
 PT - useful to screen for specific inhibitors, e.g. to search for, or
 PT study agent for treatment of various diseases
 PS Claim 2; Pages 11-12; 16pp; Japanese.
 CC The present sequence encodes the human colon carcinoma COLO 201
 CC cell line derived serine protease 60 (SP60), which can be used to
 CC screen for specific inhibitors, e.g. to search for, or study an
 CC agent for the treatment of various diseases.
 SO Sequence 699 BP; 157 A; 190 C; 209 G; 143 T;

Query Match 7.9%; Score 137; DB 1; Length 699;
 Best Local Similarity 53.7%; Pred. No. 5.4e-29;
 Matches 381; Conservative 0; Mismatches 310; Indels 18; Gaps 4;

QY 878 tctgtggcgagagagcggtcccggggctgtgcccctgagctcagcctgcagctcc 937
 DB 2 TGTGTGGGTGGGAGGAGGCGCTGTGTGGATTCTTGGCTTGGCGAGCTCAGCATCCAGTAGC 61
 QY 938 agaaagctcagcgtgtgagaggtctcatcatcaaccccgagtgatcgtgacagcgccc 997
 DB 62 ACAAAACAGCACGCTGTGTGGAGGAGCATCTGACCCCACTGGTCTCAGCGACGCC 121
 QY 998 actgctgtgaaanaaccttaacaatcatalgcatgagcagcattgtcggtgatttga 1057
 DB 122 ACTGCTTCAGGAACAT-----ACCGATGTGTTCAACTGGAAGGTGGGGCAGGCTCA 174
 QY 1058 gaaacatttcattgtctctatgagcgcgataccaagtagaanaagtattctcatccaa 1117
 DB 175 GACAAACTGGCAGCTTCCCATCCCTGCGCTGTGCGCAAGATCATCATTTGATTCAAC 234
 QY 1118 attatgacctccaagaagcaaaatagatattgctgtatgaaagctgcagagccttga 1177
 DB 235 CCCATTGTACC-----CCAAAGACAATGATGATGCGCTCTGTAAGGTGCAAGTTCCTCACTGA 289
 QY 1178 ctctcaagcactagtgaaacagatgtgtctgcccacaaccagacatgacgtctcagcag 1237
 DB 290 CTTTCTCAGGACACAGTCAAGCCCATCTGTCTGCTTCTTTATGAGGAGCTCAGCTCAG 349
 QY 1238 aacagctctctgtatcttcctgtgtggtgggagcagcagagagaa---aggaaagacctcag 1294
 DB 350 CCACCCCACTCTGATCATTTGGATGGGGCTTACGAAGACAGATGAGGAGGAGATGTCTG 409
 QY 1295 aagctgtgaaagctgcagaaagtgctctcatgtgagacaagagatgcacaagagatag 1354
 DB 410 ACATACTGCTGAGGCGGTGATCCAGGTCAATTGACAGCACACGCGCAATGACGAGACGAT 469
 QY 1355 tctatgacaactgtatcacacagcagcatgtctgtcgcgcttcctgcaggggaaagctcg 1414

DB 470 CGTACGAGGGGAGTAGTCACCGAAGATGATGTGTACAGGATCCCGAAGGGCTGTGG 529
 QY 1415 attcttgcagagtgagcagtgagggcctctgttcaacttcgaagaacaatattcgttgc 1474
 DB 530 ACACCTGCGAGGGGTGACATGTGTGGCCCTG---ATGTACCAATGTGACAGTGGCAGT 586
 QY 1475 tgataggggatacaagctgggtgtctgtgtgcgaagaacttacaagcagagatgtacg 1534
 DB 587 TGTGTGGCATCGTTAGCTGTGGGCTATGCTGCGGGGCGCCGAGACACCCAGAGATTA 646
 QY 1535 ggaatgtgtgtattatcagagcatttattcgaacaatgtggggagga 1583
 DB 647 CCAAGCTCACAGCTATCTCACTGATGTCAACATGCTGTGAGAGCTGA 695

RESULT 6
 X15134
 ID X15134 standard; cDNA; 1605 BP.
 AC X15134;
 DT 21-APR-1999 (first entry)
 DE cDNA encoding mouse serine protease called hepsin.
 KM Mouse serine protease; hepsin; animal model; bone disease;
 KM bone disorder; skeletal disorder; osteoporosis; Paget's disease;
 KM osteitis deformans; elevated bone alkaline phosphatase level; ss.
 OS Mus musculus.
 FH Key Location/Qualifiers
 FT CDS 35..1285
 FT /tag=a
 FT /product=hepsin
 FT W09854307-A1.
 PN 03-DEC-1998.
 PD 29-MAY-1998; E03199.
 PE 30-DEC-1997; US-000486.
 PR 30-MAY-1997; US-866058.
 PA (SCHD) SCHERING AG.
 PA (UNIW) UNIV WASHINGTON.
 PI Sadler JE, Wu Q;
 DR WPI: 99-070213/06.
 DR P-PSDB: W96812.

PT New nucleic acid functionally disrupts mouse hepsin gene - used to
 PT provide transgenic mice with abnormally elevated blood alkaline
 PT phosphatase, useful as models for bone disorders
 PS Claim 28; Fig 2; 29pp; English.
 CC The present sequence encodes a mouse serine protease called hepsin.
 CC The specification describes a mammalian cell in which expression of a
 CC gene encoding hepsin has been functionally interrupted or suppressed.
 CC The products and methods provide an animal model for bone disease,
 CC and are useful to determine effective treatment for bone and skeletal
 CC disorders such as osteoporosis, Paget's disease and osteitis
 CC deformans, especially those associated with elevated bone alkaline
 CC phosphatase levels.
 SO Sequence 1605 BP; 316 A; 445 C; 487 G; 357 T;

Query Match 7.4%; Score 127.8; DB 1; Length 1605;
 Best Local Similarity 50.7%; Pred. No. 3.1e-26;
 Matches 380; Conservative 0; Mismatches 342; Indels 27; Gaps 2;

QY 873 cagagattgtgcygagagcggtcccggggctgtgcccctgagcagtgactgtga 932
 DB 514 CGCATTTGGGGGGGCGAGAGACAGCTGTGGGAAGTGGCCGTGACAGTCAAGCTGCG 573
 QY 933 cgtccagaagctcagcgtgtctgagagctcatcatcaccccgagatgtatcgtacagc 992
 DB 574 TTATGATGGGACCACTCTGTGTGGGGGTCCCTGTGTGTGGGCAATGGGTGCTGACTGC 633
 QY 993 cgcacactgtcgtgaaanaaccttaacaatcatatgacattgagcgtgacttgcggggat 1052
 DB 634 TGCATTTCTTTCCAGACCGGAACGGGCTGTCTCCGTGGCGAGATTTTGTCTGTGTC 693
 QY 1053 tttagacaacttttca-----tgttctatggagcgcgataccaagttaga 1097


```

CC FT DISULFID 201 284 BY SIMILARITY.
CC FT DISULFID 227 256 BY SIMILARITY.
CC FT DISULFID 231 237 BY SIMILARITY.
CC FT DISULFID 292 375 BY SIMILARITY.
CC FT DISULFID 318 347 BY SIMILARITY.
CC FT DISULFID 322 328 BY SIMILARITY.
CC FT DISULFID 340 345 BY SIMILARITY.
CC FT DISULFID 363 503 BY SIMILARITY.
CC FT DISULFID 419 435 BY SIMILARITY.
CC FT DISULFID 517 584 BY SIMILARITY.
CC FT DISULFID 548 563 BY SIMILARITY.
CC FT DISULFID 574 602 BY SIMILARITY.
CC SEQUENCE 638 AA: 71273 MW: FEDCC211 CRC32;

Query Match 20.18; Score 740; DB 1; Length 638;
Best Local Similarity 44.58; Pred. No. 1,97e-167;
Matches 105; Conservative 45; Mismatches 78; Indels 8; Gaps 6;

DB 389 ARIYGTNSSL-GEPMQVSLQVLYSQNHMGSSITGRQMIITLAHCFD-GIPYDVMR 446
  :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 253 SRIVGG-ESALPAMPQVSLHQN---HVCSSITPEMIVTAACVKEPLNPNHMT 308
  :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 447 YGGLILSLSEITKTPSSRIKELIHOKYKSGSDIALIKIQTPLNTEFOKPLPSK 506
  :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 309 AFAGILHROSEMFYAGYQ-VEKYISHPNYDSKTKNNDIALMKLOKPLTFNDLVKPVCLPN 367
  :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 507 KAPNTIITNCWYTGWYTERGETONILQATIPLVNNECCKKR-DVYIKROMICAG 565
  :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 368 PGMMLQPEOLCWSIGWATEKGTSEVLNAKVLLETORCSRIVYDNLITPAMICAG 427
  :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 566 YKGGIDACGDSGGPLVCKHSGRMOLVGTISMGEGCARKEQPGVYTKVAEYDNI 621
  :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 428 FLOGNVDSCGDSGGPLVTSKNINMLIGDTSMGSGCAKARPGVGNVAVFTDNI 483
  :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 8
ID KAL_MOUSE STANDARD; PRT: 638 AA.
AC P26262;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE PLASMA KALLIKREIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKREIN)
DE (KININOGENIN) (FLETCHEER FACTOR).
GN KLK3 OR PK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-BALB/C; TISSUE-LIVER;
RX MEDLINE: 91090844
RA SEIDM N.G., SAWYER N., HAMELIN J., MION P., BEAUBIEN G.,
RA BRACHPAPA L., ROCHEMONT J., MBIRAY M., CHRETIEN M.;
RT "Mouse plasma kallikrein: cDNA structure, enzyme characterization,
RT and comparison of protein and mRNA levels among species.";
RL DNA Cell Biol. 9:737-748(1990).

-1- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT
CC ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING
CC TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM
CC HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
CC SYSTEM BY CONVERTING PRORENIN INTO RENIN.
CC -1- SUBUNIT: THE ZYMOGEN IS ACTIVATED BY FACTOR XIIa, WHICH CLEAVES
CC THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,
CC AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE
CC CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
CC -1- SIMILARITY: CONTAINS 4 APPLE DOMAINS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. BELONGS TO THE PLASMA KALLIKREIN SUBFAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M58588; AAA6393.1;
CC FIR: A36557; KMSPL.
CC HSSP: P00750; IRTF.
CC MGD: MGI:102849; KLK3.
CC PROSITE: PS00134; TRYPSIN_HIS; 1.
CC PROSITE: PS00135; TRYPSIN_SER; 1.
CC PROSITE: PS00495; APPLE; 4.
CC PFAM: PF00024; apple; 4.
CC PFWM: PF00089; trypsin; 1.
CC Hydrolysis: Serine protease; Glycoprotein; Plasma; Zymogen; Signal;
CC Fibrinolysis; Blood coagulation; Inflammatory response; Liver;
CC Duplication; Bradykinin.
CC SIGNAL 1 19
CC CHAIN 20 350 PLASMA KALLIKREIN HEAVY CHAIN.
CC FT CHAIN 391 638 PLASMA KALLIKREIN LIGHT CHAIN.
CC FT REPEAT 20 105 APPLE 1.
CC FT REPEAT 110 195 APPLE 2.
CC FT REPEAT 200 285 APPLE 3.
CC FT REPEAT 291 376 APPLE 4.
CC FT DOMAIN 389 621 CATALYTIC.
CC FT CARBOHYD 127 127 PROBABLE.
CC FT CARBOHYD 308 308 PROBABLE.
CC FT CARBOHYD 396 396 PROBABLE.
CC FT CARBOHYD 453 453 PROBABLE.
CC FT CARBOHYD 494 494 PROBABLE.
CC FT ACT_SITE 434 434 CHARGE RELAY SYSTEM.
CC FT ACT_SITE 483 483 CHARGE RELAY SYSTEM.
CC FT ACT_SITE 578 578 CHARGE RELAY SYSTEM.
CC FT DISULFID 21 104 BY SIMILARITY.
CC FT DISULFID 47 77 BY SIMILARITY.
CC FT DISULFID 51 57 BY SIMILARITY.
CC FT DISULFID 111 194 BY SIMILARITY.
CC FT DISULFID 137 166 BY SIMILARITY.
CC FT DISULFID 141 147 BY SIMILARITY.
CC FT DISULFID 201 284 BY SIMILARITY.
CC FT DISULFID 227 256 BY SIMILARITY.
CC FT DISULFID 231 237 BY SIMILARITY.
CC FT DISULFID 292 375 BY SIMILARITY.
CC FT DISULFID 318 347 BY SIMILARITY.
CC FT DISULFID 322 328 BY SIMILARITY.
CC FT DISULFID 340 345 BY SIMILARITY.
CC FT DISULFID 363 503 BY SIMILARITY.
CC FT DISULFID 383 503 BY SIMILARITY.
CC FT DISULFID 419 435 BY SIMILARITY.
CC FT DISULFID 517 584 BY SIMILARITY.
CC FT DISULFID 548 563 BY SIMILARITY.
CC FT DISULFID 574 602 BY SIMILARITY.
CC SEQUENCE 638 AA: 71368 MW: 8370CD2E CRC32;

Query Match 19.68; Score 720; DB 1; Length 638;
Best Local Similarity 43.48; Pred. No. 8.58e-162;
Matches 102; Conservative 44; Mismatches 83; Indels 6; Gaps 4;

DB 389 ARIYGTNSSL-GEPMQVSLQVLYSQNHMGSSITGRQMIITLAHCFD-GIPYDVMR 447
  :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 253 SRIVGGESALPAMPQVSLHQN---HVCSSITPEMIVTAACVKEPLNPNHMTA 309
  :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 448 YGGLILSLSEITKTPSSRIKELIHOKYKSGSDIALIKIQTPLNTEFOKPLPSK 507
  :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 310 FAGILRGS-EMFAGQVQYKVISHPYDSKTKNNDIALMKLOKPLTFNDLVKPVCLPNP 368
  :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 508 ADNTIITNCWYTGWYTERGETONILQATIPLVNNECCKKR-DVYIKROMICAGY 566
  :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 369 GMLQPEOLCWSIGWATEKGTSEVLNAKVLLETORCSRIVYDNLITPAMICAG 428
  :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DB 567 KEGGTADCGDSGGPLVCKHSGRMOLVGTISMGEGCARKEQPGVYTKVAEYDNI 621
  :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
QY 429 FLOGNVDSCGDSGGPLVTSKNINMLIGDTSMGSGCAKARPGVGNVAVFTDNI 483
  :||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

```

DT 16-OCT-1998 (first entry)
 DE Human factor VIIa encoding cDNA.
 KM Humun; truncated: tissue factor; tTF; TF; tumour; coagulation;
 KM blood vessel; Factor VIIa; FVIIa; benign growth; vascularised;
 KM benign prostatic hypertrophy; malignant; necrosis; angiogenesis;
 KM diabetic retinopathy; restenosis; neovascular glaucoma; psoriasis;
 KM rheumatoid arthritis; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 36..1436
 FT /tag= a "Factor VIIa"
 FT /product= "Factor VIIa"
 PN WO981394-A2.
 PD 23-JUL-1998.
 PF 20-JAN-1998; U01012.
 PR 27-MAR-1997; US-042427.
 PR 22-JAN-1997; US-035920.
 PR 27-JAN-1997; US-036205.
 PA (TEXA) UNIV TEXAS SYSTEM.
 PI Gao B, King SM, Thorpe PE;
 DR MPI: 98-413821/35.
 DR P-PSDB; W69606.
 PT Composition containing coagulation-defective tissue factor for
 PT treating, e.g. tumours - useful for, e.g. promoting coagulation in
 PT pro-thrombotic and tumour-associated vasculature, used with, e.g.
 PT factor 7 or anti-cancer agent
 PS Disclosure: Page 195; 225pp; English.
 CC A composition has been developed which comprises at least 1 coagulation-
 CC deficient tissue factor (TF) compound that is modified to increase its
 CC biological half-life, but excluding modification that involves attachment
 CC to an antibody (or its antigen-binding region) that binds to a component
 CC (cells, vasculature or stroma) of tumours. Also described in the present
 CC invention are compositions containing any coagulation-deficient TF for
 CC promoting coagulation. The coagulation-deficient TFs are used to promote
 CC coagulation preferentially in prothrombotic vessels, particularly those
 CC associated with: (i) benign growths (e.g. benign prostatic hypertrophy);
 CC (ii) vascularised, malignant tumours of medium or large size (where they
 CC also induce tumour necrosis), or (iii) other disorders that involve
 CC angiogenesis, e.g. diabetic retinopathy, restenosis, neovascular
 CC glaucoma, psoriasis and rheumatoid arthritis. The composition can be
 CC administered systemically, particularly intravenously, typically at
 CC 0.2-200 mg, given 3 times over 7 days. Truncated TF, and its variants,
 CC localise specifically in tumour-associated blood vessels after systemic
 CC administration, even though they contain no targeting agent. They cause
 CC little if any injury to normal tissue; may produce a synergistic response
 CC when used with other antitumour agents and they eliminate the multi-step,
 CC and expensive, preparation of antibody-based targeting constructs. The
 CC present sequence encodes human Factor VIIa, from the present invention.
 SO Sequence 1440 BP; 288 A; 434 C; 464 G; 254 T;

Query Match 5.5%; Score 95.4; DB 1; Length 1440;
 Best Local Similarity 47.9%; Pred. No. 3.6e-17;
 Matches 353; Conservative 0; Mismatches 366; Indels 18; Gaps 2;

QY 865 cgcacagagagattgtggcgagcgagcgctccggggcgccctggcagtc 924
 DB 660 CCCCAAGCGCGAATTGTGGGGGCAAGGTGTGCCCAAGGGGAGGTCTCATGGCGAGTC 719
 QY 925 agcctcaagtcacagacgctccagtgcgagagctcatcatcaccggcgatgc 984
 DB 720 cgtgttctgttgatagatgactcagttgtgtggggagaccctgatacaccatctgggtg 779
 QY 985 gtgaacagccacacgctcgctggaacaaaccttaacaatcatgcatgagcgaatt 1044
 DB 780 gctctccgcccgcctcactgttgcacaaatcaagaaactgagaaactgactgcgctg 839
 QY 1045 ggggggatttgagaaacattcatgctctatgagccggatatacaagtagaagaagtg 1104
 DB 840 ggcgagacacgacctcagcagcagcagcgagagagccggggggcgccaggtc 899
 QY 1105 attctcatcacaatattactcacaagacaagaacatgacatgctgctgaggaagctg 1164

DB 900 ATCATCCCGACGACGATAGTCCCGGGCCACCACCAACGACACATCGGCTGTCCGCTG 959
 QY 1165 cagaagcctctgacttcaacagcactagtaaacagtgctgtcccaa-----c 1215
 DB 960 CACACGCGCGTGTCTCTCACTGACATGTGTGTGCCCCCTGTGCGCCGACGACGCTTC 1019
 QY 1216 ccaggcatatgtctgcagacagacagcctgtctgattccgggtgtggggccacagag 1275
 DB 1020 TCTGAGAGAGAGCGTGGCGCTTCGTGCGCTTCTCATTTGGTCAAGCGGCGGCGAGCTGCTG 1079
 QY 1276 gagaagaagggaagacactcagaagtgtcgaacgctgcgaagtgctctcaltggaacag 1335
 DB 1080 GACCGTGGCGCCACGCGCCCTCGAGACTCATGCTCAACGCTCCCGCGGTGATGACACAG 1139
 QY 1336 agatgc-----aacagcagatatgtctatgaaacactgtctacacagcagatgc 1386
 DB 1140 GACTGCTGCGAGAGTCAACGGAAGGTGGAGACTCCCAAAATATCACGGAGTACATGTTCC 1199
 QY 1387 tgtgcggctctctgcaggggagcgtcgatctcttgcgaaggtagaagtgagggcctctg 1446
 DB 1200 TGTGCGCGTACTCTCGGATGACAGCAAGACTCCTGCAAGGGGACAGTGAAGGCCACAT 1259
 QY 1447 gtcaactcgaagaacataatctgtgtctgatatagggatatacaagctggggtcttgcctg 1506
 DB 1260 GCCACCCACTACCGGGGACAGCTGTAACCTGACGGGATCTGTCAGCTGGGGCCAGGGCTGC 1319
 QY 1507 gccaaagctcacagaccagagtgtagggagatgtatgtattcagagctgattat 1566
 DB 1320 GCAACCGTGGGCCACTTTGGGGGTGTACACAGGAGTCTCCACATGATACAGTGGCTGCCAA 1379
 QY 1567 cgacaatagaggcgaga 1583
 DB 1380 AAGCTCATGCGCTCAGA 1396

RESULT 15
 ID N60063 standard; cDNA: 2177 BP.
 AC N60063;
 DT 23-MAY-1991 (first entry)
 DE Partial Factor VII cDNA.
 KM Factor VII; Factor VIIa.
 FH Key Location/Qualifiers
 FT CDS 13..1128
 FT /tag= a
 PN EP-200421-A.
 PD 10-DEC-1986.
 PF 16-APR-1986; 302855.
 PR 17-APR-1985; US-724311.
 PR 16-DEC-1985; US-810002.
 PA (ZYMO-) ZYMOGENETICS INC.
 PI Hagen FS, Murry MJ, Berkner KL, Insley MY, Woodbury RG,
 PI Gray CL;
 DR MPI: 86-326899/50.
 DR P-PSDB; P60055.
 PT DNA construct used to transfect hosts - to produce protein which
 PT activates to give factor VIIa
 PS Disclosure; Fig. 1A; 55pp; English.
 CC The partial factor VII cDNA sequence is produced by joining portions
 CC of cDNA clones lambda V12115 and lambda V11923. It is used in a DNA
 CC construct which contains a nucleotide sequence encoding a protein
 CC which, on activation, has the same biological activity for blood
 CC coagulation as factor VIIa. The nucleotide codes at least partially
 CC for factor VII and comprises a sequence encoding a calcium binding
 CC domain joined to a second sequence downstream of this encoding a
 CC catalytic domain for the serine protease activity of factor VIIa. The
 CC calcium binding domain comprises a gene encoding factor VII, IX, X,
 CC protein C, prothrombin or protein S. The construct is used to
 CC transfect host cells to produce the protein which, on activation,
 CC yields factor VIIa.
 SO Sequence 2177 BP; 569 A; 624 C; 605 G; 379 T;

FT	DOMAIN	106	492	EXTRACELLULAR (POTENTIAL).
FT	DOMAIN	111	150	LDI-RECEPTOR CLASS A.
FT	DOMAIN	151	243	SRCR.
FT	ACT_SITE	256	492	CATALYTIC.
FT	ACT_SITE	296	296	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	345	345	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	441	441	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	DISULFID	113	126	BY SIMILARITY.
FT	DISULFID	120	139	BY SIMILARITY.
FT	DISULFID	133	148	BY SIMILARITY.
FT	DISULFID	244	365	BY SIMILARITY.
FT	DISULFID	281	297	BY SIMILARITY.
FT	DISULFID	410	426	BY SIMILARITY.
FT	DISULFID	437	465	BY SIMILARITY.
FT	CARBOHYD	213	213	POTENTIAL.
FT	CARBOHYD	249	249	POTENTIAL.
SC	SEQUENCE	492 AA;	53647 MM; 19BC1F67 CRC32;	
Query Match				
Best Local Similarity		99.0%;	Score 3635; DB 1; Length 492;	
Matches		485; Conservative	6; Mismatches 0; Indels 1; Gaps 1	
Db	1 MALNSGSPRALGPIYENHNGIYOPENPYPAQPIVTVVYVHHPAQYTPSPVPQYAPRVLQA	60		
Qy	1 MALNSGSPRALGPIYENHNGIYOPENPYPAQPIVTVVYVHHPAQYTPSPVPQYAPRVLQA	60		
Db	61 SNPVYCTOPKSPSGVCTSKTKKALCITLTGTFELYGAAALAGLLMKFMSKCSNSGIEC	120		
Qy	61 SNPVYCTOPKSPSGVCTSKTKKALCITLTGTFELYGAAALAGLLMKFMSKCSNSGIEC	120		
Db	121 DSSGTCINPSMWCQGVSHCPGEGEDENRCVRLYGNPFIQMTYSORKSWHPYCODDMENY	180		
Qy	121 DSSGTCINPSMWCQGVSHCPGEGEDENRCVRLYGNPFIQMTYSORKSWHPYCODDMENY	180		
Db	181 GRACRQWGYKNPNFSSQGIYDDSGSTSEFMKLNISAGVNDIYKRLYHSDACSSRAVYSLR	240		
Qy	181 GRACRQWGYKNPNFSSQGIYDDSGSTSEFMKLNISAGVNDIYKRLYHSDACSSRAVYSLR	239		
Db	241 CIACGVNLNSSRQSHIVGEGESALPGAMPQVSLYHQNVAHVGSGSITTPEWITVAHCVEK	300		
Qy	240 CIACGVNLNSSRQSHIVGEGESALPGAMPQVSLYHQNVAHVGSGSITTPEWITVAHCVEK	299		
Db	301 PLNNPMHTATAGILIRQSFMYGAGIYQVQKXISHPNYDSKTKKNDIALMKLOKPLTFENDL	360		
Qy	300 PLNNPMHTATAGILIRQSFMYGAGIYQVQKXISHPNYDSKTKKNDIALMKLOKPLTFENDL	359		
Db	361 VKPVCILPMPGMMLOPEQICWISGWCATEEKKTSSEVLNAAVLLIETORCSRYVYDMLI	420		
Qy	360 VKPVCILPMPGMMLOPEQICWISGWCATEEKKTSSEVLNAAVLLIETORCSRYVYDMLI	419		
Db	421 TPAMICAGFLOGNDVSCOGSGGGLVTSNNIMWLIGTSMGSGCAKAYRPGYGVNVAVF	480		
Qy	420 TPAMICAGFLOGNDVSCOGSGGGLVTSNNIMWLIGTSMGSGCAKAYRPGYGVNVAVF	479		
Db	481 TDMITRQMKANG	492		
Qy	480 TDMITRQMKANG	491		
RESULT				
ID	ENTR	2	STANDARD;	PRT; 1035 AA.
AC	P86072;			
DT	01-FEB-1996 (Rel. 33, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	15-DEC-1999 (Rel. 39, Last annotation update)			
DE	ENTEROPEPTIDASE PRECURSOR (EC 3.4.21.9) (ENTEROKINASE).			
GN	PRSS7 OR ENTK.			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
OC	Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;			
OC	Bovina; Bos.			
RN	(1)			
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.				

RC	TISSUE-DUODENUM;
RX	MEDLINE; 94329561.
RA	KIRAMOTO Y., YUAN X., WU O., MCCOURT D.W., SADDLER J.E.;
RT	"ENTEROKINASE, THE INITIATOR OF INTESTINAL DIGESTION, IS A MOSAIC
RL	PROTEASE COMPOSED OF A DISTINCTIVE ASSORTMENT OF DOMAINS.";
RN	Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
RP	[2]
RX	SEQUENCE OF 801-1035 FROM N.A., AND PARTIAL SEQUENCE.
RA	MEDLINE; 94043122.
RT	LAVALLE E.R., REHEMUTULLA A., RACIE L.A., DIBLASIO E.A.,
RL	PRENEN C., GRANT K.L., LIGHT A., MCCOY J.M.;
RN	"CLONING AND FUNCTIONAL EXPRESSION OF A CDNA ENCODING THE CATALYTIC
RP	SUBUNIT OF BOVINE ENTEROKINASE.";
RX	J. Biol. Chem. 268:23311-23317(1993).
RC	[3]
RA	SEQUENCE OF 801-827.
RT	TISSUE-INTESTINE;
RL	MEDLINE; 92189715.
RA	LIGHT A., JANSKA H.;
RT	"THE AMINO-TERMINAL SEQUENCE OF THE CATALYTIC SUBUNIT OF BOVINE
RL	ENTEROKINASE.";
RN	J. Protein Chem. 10:475-480(1991).
RP	-1 FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
CC	PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
CC	A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
CC	TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
CC	PROCARBOXYPEPTIDASES, AND PROELASTASES.
CC	-1 CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
CC	TRYPSINOGEN.
CC	-1 SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
CC	MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.
CC	-1 SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC	-1 TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.
CC	-1 PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
CC	CLEAVED BY A TRYPSIN-LIKE PROTEASE.
CC	-1 SIMILARITY: CONTAINS 2 LDB-RECEPTOR CLASS A DOMAINS.
CC	-1 SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC	-1 SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC	-1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC	TRYPSIN FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see http://www.isb-sib.ch/announcement/).
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; U09859; AAB40026.1; -
DR	EMBL; LI9663; AAA16035.1; -
DR	PIR; A61436; A61436.
DR	HSSP; P00763; IDPO.
DR	PROSITE; PS00134; TRYPSIN_HIS; 1.
DR	PROSITE; PS01180; TRYPSIN_SER; 1.
DR	PROSITE; PS00740; MAM_1; 1.
DR	PROSITE; PS50060; MAM_2; 1.
DR	PROSITE; PS01209; LIDRA_1; 2.
DR	PROSITE; PS50068; LIDRA_2; 2.
DR	PFAM; PF00057; ldl_recept_a; 2.
DR	PFAM; PF00089; trypsin; 1.
DR	PFAM; PF00431; CUB; 2.
DR	PFAM; PF00530; SRCR; 1.
DR	PFAM; PF00629; MAM; 1.
DR	PFAM; PF01390; SEA; 1.
KW	Signal.anchor; Glycoprotein; Myristate; Hydrolase;
KW	Signal.anchor; Lysomogen; Transmembrane; Repeat; Alternative splicing.
FT	CHAIN 1 800 NON-CATALYTIC CHAIN (HEAVY CHAIN).
FT	CHAIN 801 1035 CATALYTIC CHAIN (LIGHT CHAIN).
FT	CHAIN 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT	TRANSMEM 19 47 SIGNAL-RECEPTOR CLASS A 1.
FT	DOMAIN 197 238

HSX1-4 400001 510000
HSX1-5 500001 610000
HSX1-6 600001 613769
Continuation (6 of 7) of HSX1 from base 500001 (AJ011929 Homo sapiens chromosome 21 clones, 1/1999)

Query Match 10.2% Score 176.6; DB 32; Length 110000;
Best Local Similarity 93.0%; Pred. No. 1.7e-36;
Matches 185; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 988 acagccgccacgctgctggaacaccccttaacatccatggcattgagcgattcg 1047
DB 92255 ACTGAGTCTTCTCTCCAGACCTCTTAACATCCATGCGATGCGGCGATTGCG 92196
QY 1048 gggatttggacaacattctatgtctatggagcgagatccaaagtaaaagtatt 1107
DB 92195 GGGATTTTGGACAACTTTTATGATGAGCCGGATACCAAGTAAAGTATGAT 92136
QY 1108 tctatccaattatgactccaagacaagaacaaatgacattgcgtatgaagctgag 1167
DB 92135 TCTCTCCAAATTATGATCTCCAGACCAAGAACATGACATTGCCCTGATGAGCTGCAG 92076
QY 1168 aagcctctgacttccaag 1186
DB 92075 AAGCCTCTGACTTTCACG 92057

RESULT 5
HUM8DC212 3447 bp DNA PRI 22-AUG-1994
LOCUS Homo sapiens (subclone H8 2_e7) from P1 35 H5 C8) DNA sequence.
ACCESSION L35676
VERSION L35676.1 GI:532019
KEYWORDS MX1 region.
SOURCE Homo sapiens (library: Subclones in pSP72 from P1 clone 35 H5 C8 (H8)) DNA.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 3447)
AUTHORS Martin, C.H., Bondoc, M.M., Chiang, A., Cloutier, T., Davis, C.A.,
Ericsson, C.L., Jaklevic, M.A., Kim, R.J., Lee, M.T., Li, M.,
Mayeda, C.A., Stelert-Ei Kheir, A. and Palazolo, M.J.,
Sequencing of the MX1 region on human chromosome 21
Unpublished (1994)
COMMENT Sequence submitted by:
Human Genome Center and
Drosophila Genome Center
Lawrence Berkeley Laboratory
Berkeley, CA 94720
e-mail: seq@genome.lbl.gov
This subclone overlaps H8 3_b5 and H8 6_e1.

FEATURES
source 1. 3447
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_lib="Subclones in pSP72 from P1 clone 35 H5 C8 (H8)"
Location/Qualifiers

BASE COUNT 804 a 790 c 1024 g 829 t
ORIGIN

Query Match 9.1%; Score 157.4; DB 9; Length 3447;
Best Local Similarity 99.4%; Pred. No. 1.6e-31;
Matches 158; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1423 cagggatgaagcggggggcctctgtctacattggaagaacataatcggcgtgataagg 1482
DB 2290 CAGGGTGAAGTGGAGGGCCTCTGTGCTTGAAGAACAAATTCGTGGTGATAGGG 2349
QY 1483 gatcaagctggggtctctgctgctgccaagcttacagacgaagatgtaagggaattg 1542

DB 2350 GATACACCTGGGGTTCTGCTGTCCCAAGCTTACAGACCAAGAGTGTACGGGATGTG 2409
QY 1543 atggtatcacgagcaggtattatcagacaatgaggga 1581
DB 2410 ATGGTATCAGGAGTGGATTTATGCAAAATGAGGGA 2448

RESULT 6
HUM8DC202 2330 bp DNA PRI 22-AUG-1994
LOCUS Homo sapiens (subclone H8 3_b5 from P1 35 H5 C8) DNA sequence.
ACCESSION L35675
VERSION L35675.1 GI:532018
KEYWORDS MX1 region.
SOURCE Homo sapiens (library: Subclones in pSP72 from P1 clone 35 H5 C8 (H8)) DNA.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 2330)
AUTHORS Martin, C.H., Bondoc, M.M., Chiang, A., Cloutier, T., Davis, C.A.,
Ericsson, C.L., Jaklevic, M.A., Kim, R.J., Lee, M.T., Li, M.,
Mayeda, C.A., Stelert-Ei Kheir, A. and Palazolo, M.J.,
Sequencing of the MX1 region on human chromosome 21
Unpublished (1994)
COMMENT Sequence submitted by:
Human Genome Center and
Drosophila Genome Center
Lawrence Berkeley Laboratory
Berkeley, CA 94720
e-mail: seq@genome.lbl.gov
This subclone overlaps H8 4_d4 and H8 2_e7.

FEATURES
source 1. 2330
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_lib="Subclones in pSP72 from P1 clone 35 H5 C8 (H8)"
Location/Qualifiers

BASE COUNT 501 a 578 c 666 g 585 t
ORIGIN

Query Match 8.4%; Score 146; DB 9; Length 2330;
Best Local Similarity 100.0%; Pred. No. 1.7e-28;
Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1578 ggcagacggctaatccacatgctctgctccttgacgtgtttcaagaagaacatgg 1637
DB 981 GGCAGACGGCTAATCCACATGCTCTGCGCTTGACGTCGTTTACMAAAACATGGG 1040
QY 1638 gctgttttgcctcccgcgatgatttactcttagagatgattcagaggtacttcatt 1697
DB 1041 GCTGTTTGCCTCCCGCGCATGATTACTCTTAGAGTGAATTCAGAGGTCACTTCATT 1100

QY 1698 ttattaaacagtgaactgtctggc 1723
DB 1101 TTTATTAACAGTGAAGTGTCTGGC 1126

RESULT 7
HUM8DC222 3599 bp DNA PRI 22-AUG-1994
LOCUS Homo sapiens (subclone H8 6_e1 from P1 35 H5 C8) DNA sequence.
ACCESSION L35677
VERSION L35677.1 GI:532020
KEYWORDS MX1 region.
SOURCE Homo sapiens (library: Subclones in pSP72 from P1 clone 35 H5 C8 (H8)) DNA.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 3599)
AUTHORS Martin, C.H., Bondoc, M.M., Chiang, A., Cloutier, T., Davis, C.A.,

```

FT CHAIN 1 829 NON-CATALYTIC CHAIN (HEAVY CHAIN).
FT CHAIN 830 1069 CATALYTIC CHAIN (LIGHT CHAIN).
FT TRANSMEM 19 47 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT DOMAIN 227 268 LDL-RECEPTOR CLASS A 1.
FT DOMAIN 270 379 CUB.
FT DOMAIN 387 549 MAM.
FT DOMAIN 569 679 CUB.
FT DOMAIN 686 724 LDL-RECEPTOR CLASS A 2.
FT DOMAIN 723 816 SRCR.
FT ACT_SITE 874 874 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 925 925 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 1021 1021 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT LITID 2 2 MRISTATE (POTENTIAL).
FT DISULFID 229 242 BY SIMILARITY.
FT DISULFID 236 255 BY SIMILARITY.
FT DISULFID 249 266 BY SIMILARITY.
FT DISULFID 688 700 BY SIMILARITY.
FT DISULFID 695 713 BY SIMILARITY.
FT DISULFID 707 722 BY SIMILARITY.
FT DISULFID 817 945 INTERCHAIN (BY SIMILARITY).
FT DISULFID 859 1027 BY SIMILARITY.
FT DISULFID 959 1027 BY SIMILARITY.
FT DISULFID 991 1006 BY SIMILARITY.
FT DISULFID 1017 1045 BY SIMILARITY.
FT CARBOHYD 147 147 POTENTIAL.
FT CARBOHYD 197 197 POTENTIAL.
FT CARBOHYD 212 212 POTENTIAL.
FT CARBOHYD 373 373 POTENTIAL.
FT CARBOHYD 380 380 POTENTIAL.
FT CARBOHYD 433 433 POTENTIAL.
FT CARBOHYD 515 515 POTENTIAL.
FT CARBOHYD 579 579 POTENTIAL.
FT CARBOHYD 675 675 POTENTIAL.
FT CARBOHYD 727 727 POTENTIAL.
FT CARBOHYD 751 751 POTENTIAL.
FT CARBOHYD 770 770 POTENTIAL.
FT CARBOHYD 791 791 POTENTIAL.
FT CARBOHYD 897 897 POTENTIAL.
FT CARBOHYD 936 936 POTENTIAL.
FT CARBOHYD 999 999 POTENTIAL.
SQ SEQUENCE 1069 AA; 118735 MW; 488825A2 CRC32;

Query Match 23.08; Score 845; DB 1; Length 1069;
Best Local Similarity 36.28; Pred. No. 2,73e-197;
Matches 139; Conservative 87; Mismatches 138; Indels 20; Gaps 17;

```

```

RESULT 6
ID KAL_HUMAN STANDARD: PRT; 638 AA.
AC P03952;
DT 23-OCT-1986 (Rel. 02, Created)
DT 23-OCT-1986 (Rel. 02, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE PLASMA KALLIKREIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKREIN)
DE (KININOGENIN) (FLETCHER FACTOR).
OS KLK3.
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE; 86243359.
RA CHUNG D W, FUJIKAWA K, MCMULLEN B A, DAVIE E W;
RT "Human plasma prekallikrein, a zymogen to a serine protease that
RT contains four tandem repeats."
RL Biochemistry 25:2410-2417(1986).
[2]
RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.
RX MEDLINE; 91152016.
RA MCMULLEN B A, FUJIKAWA K, DAVIE E W;
RT "Location of the disulfide bonds in human plasma prekallikrein: the
RT presence of four novel apple domains in the amino-terminal portion of
RT the molecule."
RL Biochemistry 30:2050-2056(1991).
-1- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT
-1- ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING
-1- TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM
-1- HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
-1- SYSTEM BY CONVERTING PRORENIN INTO RENIN.
-1- SUBUNIT: THE ZYMOGEN IS ACTIVATED BY FACTOR XIIA, WHICH CLEAVES
-1- THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,
-1- AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE
-1- CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
-1- DISEASE: DEFECTS IN KLK3 ARE THE CAUSE OF FLETCHER FACTOR
-1- DEFICIENCY, A BLOOD COAGULATION DEFECT.
-1- SIMILARITY: CONTAINS 4 APPLE DOMAINS.
-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
-1- TRYPSIN FAMILY. BELONGS TO THE PLASMA KALLIKREIN SUBFAMILY.
CC -----
CC EMBL; M13143; AAA60153.1; -.
CC PIR; A00921; KOHRP.
CC PIR; A57939; A37939.
CC HSP; P00763; IDPO.
CC MIM; 229000; -.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
CC PROSITE; PS00495; APPLE; 4.
CC PFAM; PF00024; Apple; 4.
CC PFAM; PF00089; trypsin; 1.
CC F00089; Serine protease; Glycoprotein; Plasma; Zymogen; Signal;
CC Fibrinolysis; Blood coagulation; Inflammatory response; Liver;
CC Duplication; Bradykinin.
FT SIGNAL 1 19
FT CHAIN 20 390 PLASMA KALLIKREIN HEAVY CHAIN.
FT CHAIN 391 638 PLASMA KALLIKREIN LIGHT CHAIN.
FT REPEAT 20 105 APPLE 1.
FT REPEAT 110 195 APPLE 2.
FT REPEAT 200 285 APPLE 3.
FT REPEAT 291 376 APPLE 4.
FT DOMAIN 389 621 CATALYTIC.

```


[illegible]

Db	984	CCACTCTCTATGATCTCCCTGACCTCTACAGAAATACATCATCAGACCACTGATCTGCTGCTGC	1043
QY	1218	aggcattgctctcagccagcaaacgctctctctgattccgggtgaggggccacagggga	1277
Db	1044	GGGACAGAGCCCTGGTGGATGTCAGAGGTCTGTACGTGACCGCGTGGGGTAAACACAGTT	1103
QY	1278	gaagggaagagccctcagaagtgctggaagcgcgcgaaggtgtcttcatttgacacagag	1337
Db	1104	CTATAGGCCAACAGGCTATGTGGTCTCCAGAGAGCCCGGGTTCCATTCATTAAGCAACGAAGT	1163
QY	1338	atgcacagacagaaatgctctatgacaaacctgatcacacccagcaatgactctgtccggctt	1397
Db	1164	TGTGAACAGCCCGACACTTCTACGGGGAATCAATCAAGCCCAAGATGTCTGTGCTGGCTA	1223
QY	1398	ctctgagggaagcgtcgattctctgcagggtgacagtgaggaagggcctctgtc-----	1449
Db	1224	TCCGAGAGGTGGCATTTGTCGCTGCCAGGGCGACAGTGAGAGCCCTTTGTGTGAAGA	1283
QY	1449	---cacttcgaagaacaatactctgtgtgtgtctgataaggatcaagcttgagggtcttgagtg	1505
Db	1284	CAGATCTCTGGGACATCAAGGTGGCGGCTATGTGGCATTTTAAGCTGGGGTACCGGCTG	1343
QY	1506	tgcacaaagcttacagcagcagagtgtaeaggaaatgtatgtatcaccgagctgattca	1565
Db	1344	TGCTTTGGCCCCGGAAGCCAGAGGTGTACACCAAGTCACTCACTTCGGGAGTGATCTT	1403
QY	1566	tcgacaaatgaggggcagacggcctaataca	1594
Db	1404	CAAGGCCATTAAGACTCACTCCGAAGCCA	1432
RESULT	14		
LOCUS	AB022425		
DEFINITION	Sus scrofa mRNA	MM	21-JAN-1999
ACCESSION	AB022425		
VERSION	AB022425.1	GI:4165314	
KEYWORDS	Kallikrein.		
SOURCE	Sus scrofa plasma CDNA to mRNA.		
ORGANISM	Sus scrofa		
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Artiodactyla; Suidae; Sus; Suidae; Sus.		
AUTHORS	1 (bases 1 to 2230)		
TITLE	Takahashi, T., Kimura, A., Okimura, H. and Hamabata, T.		
JOURNAL	Submitted (07-JAN-1999) to the DDBJ/EMBL/Genbank databases. Direct Submission		
REFERENCE	Submitted (07-JAN-1999) to the DDBJ/EMBL/Genbank databases. Takayuki Takahashi, Hokkaido University, Graduate School of Science, Kitaku Kita 10 Joh Nishi 8 chome, Sapporo, Hokkaido 060-0810, Japan (E-mail: ttakahase@ci.hokudai.ac.jp, Tel: 81-11-706-2748, Fax: 81-11-706-2748)		
AUTHORS	2 (bases 1 to 2230)		
TITLE	Takahashi, T., Kimura, A., Okimura, H. and Hamabata, T.		
JOURNAL	Porcine liver plasma kallikrein		
FEATURES	Published Only in Database (1999) In press		
SOURCE	Location/Qualifiers		
	1..2230		
	/organism="Sus scrofa"		
	/db_xref="taxon:9823"		
	/tissue_type="plasma"		
	14..1945		
	/codon_start=1		
	/product="kallikrein"		
	/protein_id="BA37147.1"		
	/db_xref="GI:4165315"		
	/translation="MEVILFRISFROAVFMCLFAVSGCLLPOLHKNTFFRGSDV		
	SAWTPSARHOMMCTPRLCLSPFLPADSTVDRKCEFLDVTGMLPVLRENEHLCFFY		
	AISHSLKOCCHQIRACHRDYKRIIDRGVAVNSKYTVEECRCNSHLCFFY		
	ATQAFNNKEIKNCLKHSPEGRTITSPHCLFFETYNAMITISORVCELUKTSHGTPSE		
	PSDVAIVAPDAFVFCCTITCPHCLFFETYNAMITISORVCELUKTSHGTPSE		
	PTPEDNAISGYSLTCKOTLPCHSKYISVEDEGEIANTFVQGANLCECTKTYI		
	RCQFFYSLHDEPDRGKSCILSDSGSPFKTHGMRASGYSLRCRSGDSACAP		
	TKANRYTIGGDSFLDEMPQVSOAKRLRAONHLCGSIIGHQVTLAHCFDGLSLP		
	DKMLYGLINLISETKTPESQVEIITHNYNVLDSGHIALLKLETPANTYDFQK		

M P E S R L H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPearch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Apr 19 23:44:38 2000; Maspar time 9.00 Seconds
Tabular output not generated. 706.916 Million cell updates/sec

Title: >US-09-323-597-2
Description: (1-491) from US09323597.pep
Perfect Score: 3673
Sequence: 1 MALNSGSPALGPYENHGY.....VYGNVMTDTWYRMDAG 491

Scoring table:
PAM 150
Gap 11

Searched: 131253 seqs, 12956647 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-Issued
1:5A.COMB 2:5B.COMB 3:PCN9.COMB 4:Backfiles1

Statistics: Mean 33.500; Variance 148.979; scale 0.225

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	895	24.4	798	1	US-08-200-Sequence 2, Applicatio	3.00e-71
2	895	24.4	798	3	PCT-US94-0 Sequence 2, Applicatio	3.00e-71
3	895	20.1	638	2	US-08-681-Sequence 3, Applicatio	1.19e-56
4	697	19.0	418	1	US-08-508-Sequence 25, Applicati	1.28e-52
5	691	18.8	356	2	US-08-681-Sequence 1, Applicatio	4.67e-52
6	692	18.8	855	2	US-09-027-Sequence 2, Applicatio	3.76e-52
7	678	18.5	232	1	US-08-508-Sequence 19, Applicati	7.68e-51
8	672	18.3	256	2	US-09-027-Sequence 3, Applicatio	2.79e-50
9	668	18.2	416	2	US-09-000-Sequence 2, Applicatio	6.61e-50
10	634	17.3	812	1	US-08-326-Sequence 1, Applicatio	9.85e-47
11	634	17.3	812	1	US-08-248-Sequence 1, Applicatio	9.85e-47
12	634	17.3	812	2	US-08-429-Sequence 1, Applicatio	9.85e-47
13	634	17.3	812	1	US-08-452-Sequence 1, Applicatio	9.85e-47
14	634	17.3	812	2	US-08-866-Sequence 1, Applicatio	9.85e-47
15	634	17.3	812	1	PCT-US95-0 Sequence 1, Applicatio	9.85e-47
16	634	17.3	812	1	US-08-451-Sequence 1, Applicatio	9.85e-47
17	634	17.3	812	1	US-08-605-Sequence 1, Applicatio	9.85e-47
18	634	17.3	812	2	US-08-612-Sequence 1, Applicatio	9.85e-47
19	607	16.5	270	2	US-08-978-Sequence 8, Applicatio	3.22e-44
20	598	16.3	346	4	5200340-6 Patent No. 5200340.	2.21e-43
21	597	16.3	790	1	US-08-469-Sequence 54, Applicati	2.74e-43
22	597	16.3	790	2	US-08-469-Sequence 54, Applicati	2.74e-43
23	597	16.3	791	1	US-08-643-Sequence 1, Applicatio	2.74e-43

24	597	16.3	791	2	US-08-832-Sequence 1, Applicatio	2.74e-43
25	597	16.3	791	2	US-09-131-Sequence 1, Applicatio	2.74e-43
26	597	16.3	810	4	5200340-8 Patent No. 5200340.	2.74e-43
27	597	16.3	810	1	US-07-854-Sequence 2, Applicatio	2.74e-43
28	597	16.3	810	1	US-08-147-Sequence 29, Applicatio	6.44e-43
29	593	16.1	230	2	US-08-889-Sequence 2, Applicatio	6.44e-43
30	593	16.1	230	1	US-08-379-Sequence 2, Applicatio	6.44e-43
31	593	16.1	230	1	US-08-147-Sequence 2, Applicatio	6.44e-43
32	590	16.1	814	1	US-08-750-Sequence 1, Applicatio	1.22e-42
33	587	16.0	273	2	US-08-978-Sequence 6, Applicatio	2.32e-42
34	585	15.9	230	1	US-08-456-Sequence 47, Applicati	3.57e-42
35	585	15.9	230	2	US-08-892-Sequence 47, Applicati	3.57e-42
36	585	15.9	230	1	US-08-266-Sequence 47, Applicati	3.57e-42
37	585	15.9	230	2	US-08-766-Sequence 12, Applicati	3.57e-42
38	582	15.8	273	2	US-08-978-Sequence 3, Applicatio	6.77e-42
39	577	15.7	276	2	US-09-016-Sequence 15, Applicati	1.97e-41
40	577	15.7	276	2	US-08-978-Sequence 21, Applicati	1.97e-41
41	572	15.6	267	2	US-09-016-Sequence 23, Applicati	5.73e-41
42	572	15.6	267	2	US-08-978-Sequence 18, Applicati	5.73e-41
43	572	15.6	273	2	US-08-978-Sequence 14, Applicati	5.73e-41
44	572	15.6	274	2	US-08-978-Sequence 16, Applicati	5.73e-41
45	572	15.6	274	2	US-09-016-Sequence 21, Applicati	5.73e-41

ALIGNMENTS

RESULT 1
ID US-08-200-900A-2 STANDARD: PRT: 798 AA.

XX xxxxxx

XX Sequence 2, Application US/08200900A

CC Sequence 2, Application US/08200900A

CC Patent No. 5665566

CC GENERAL INFORMATION:

CC APPLICANT:

CC TITLE OF INVENTION: CLONING OF ENTEROKINASE AND METHOD OF USE

CC NUMBER OF SEQUENCES: 38

CC CORRESPONDENCE ADDRESSES:

CC ADDRESSER: Genetics Institute, Inc. - Legal Affairs

CC STREET: 87 Cambridgepark Drive

CC CITY: Cambridge

CC STATE: MA

CC COUNTRY: USA

CC ZIP: 02140

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patent in Release #1.0, Version #1.25 (EPO)

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/200,900A

CC FILING DATE: 23-FEB-1994

CC CLASSIFICATION: 435

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Mainert, Maureen C.

CC REGISTRATION NUMBER: 31,544

CC REFERENCE/DOCKET NUMBER: GI 5201-FWC

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (617) 876-1170 x6574

CC TELEFAX: (617) 876-5851

CC INFORMATION FOR SEQ ID NO: 2:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 798 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE 798 AA; 88937 MW; 3292434 CN;

DB	157	RPCLPASHHEFEGLHOMINGKALREGGPISNALOKVDYQULIPOLCGSEVRYO--VT	214
OY	361	KPVCLPPNGMMLQEPOLCWSISGWATEKEKGTSTVLNAAKVLLIETORCSRYVDLIT	420
Db	215	PRMLCAGRKCKKACGDSGGLPVCKALSGRNFVLGVSGLCGRPNTEGYTRITGV	274
OY	421	PAMICAGLQGNVDSCGDSGGLPV-7SKNNIMVLLIDTSMGSCCAKAYRGVGNMVF	479
Db	275	ISWI 278	
OY	480	TDWI 483	
RESULT	5		
ID	060235	PRELIMINARY:	PRT: 418 AA.
AC	060235:		
DT	01-AUG-1998 (TREMBlrel. 07, Created)		
DT	01-AUG-1998 (TREMBlrel. 07, Last sequence update)		
DT	01-NOV-1999 (TREMBlrel. 12, Last annotation update)		
DE	ALWAY TRYPSIN-LIKE PROTEASE.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
OC	Eutheria; Primates; Catarrhini; Hominoidea; Homo.		
RP	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE: 9825382.		
RA	YAMAOKA K., MASUDA K., OGAWA H., TAKAGI K., UMEMOTO N., YASUOKA S.;		
RT	"Cloning and characterization of the cDNA for human alway trypsin-		
RT	like protease."		
RL	J. Biol. Chem. 273:11895-11901(1998).		
DR	EMBL: AB002134; BAA28691.1; .		
DR	HSSP: P00750; 1RPF.		
DR	PFAM: PF00390; SEA: 1.		
DR	PFAM: PF00089; trypsin: 1.		
DR	PRINTS: PR00722; CHYMOTRYPSIN.		
DR	Protease.		
SO	SEQUENCE	418 AA; 46263 MW; 0A0E0421 CRC32;	
Query Match	19.0%;	Score 697;	DB 4; Length 418;
Best Local Similarity	42.3%;	Fred. No. 4.04e-153;	
Matches 105;	Conservative 49;	Mismatches 87;	Indels 7; Gaps 6;
Db	173	CGA--CPDLITLSEORILCTEAEESPMQVSLRLNNAHCGGSLINNMWLTAAHCF-R	230
OY	240	CIACGVNINSSROSTIVGECESALPRAMPQVSLHYQNVHCGGSLITPEWVIAHCEK	299
Db	231	SNSNRDVIATSGI-STTFP--KLIMRYNLIINHNNKSAATHENDIALVRLSNSVFTKD	287
OY	300	PLNNMWHMTAAAGILROSFMEYAGAYEKEYISHPNYDSKTRNDIALMKLOKPLTFENDL	359
Db	268	IHSVCLPAAATONIPGSRATVYTGCAOXYAAGHTYBELROGOVRIISNDVCAPHSYNGAI	347
OY	360	VKPVCLPPNGMMLQEPOLCWSISGWATEKEKGTSEVLNAAVVLLIETORCSRYVDNLI	419
Db	348	LSGMLCAGVPQGVADACGDSGGLPVQDSERLMEFIVIGVSMGDCQGLPDRPGVYTRYTA	407
OY	420	TPAMICAGFLQGNVDSCGDSGGLPVYTSKNN-IMVLLIGDTSMGSGCAKAYRGVGNMVF	478
Db	408	YLDWI-RQ 414	
OY	479	FTDWIRQ 486	
RESULT	6		
ID	091495	PRELIMINARY:	PRT: 571 AA.
AC	091495:		
DT	01-NOV-1999 (TREMBlrel. 12, Created)		
DT	01-NOV-1999 (TREMBlrel. 12, Last sequence update)		
DT	01-NOV-1999 (TREMBlrel. 12, Last annotation update)		
DE	PLATELET FACTOR XI.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		

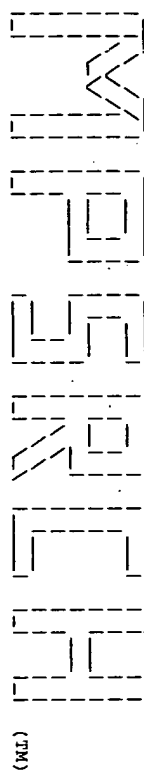
[illegible]

CC ATTORNEY/AGENT INFORMATION:
CC NAME: Johnson, James D.
CC REGISTRATION NUMBER: 31,771
CC REFERENCE/DOCKET NUMBER: 05213-0122
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 404-818-3700
CC TELEFAX: 404-818-3799
CC INFORMATION FOR SEQ ID NO: 1:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 812 amino acids
CC TYPE: amino acid
CC STRANDEDNESS:
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: NO
CC ORIGINAL SOURCE:
CC ORGANISM: Murine
CC
SQ SEQUENCE 812 AA: 90846 MW: 3411937 CN;

Query Match 17.3%; Score 634; DB 3; Length 812;
Best Local Similarity 41.4%; Pred. No. 9,85e-47;
Matches 111; Conservative 51; Mismatches 93; Indels 13; Gaps 12;

Db 553 DYCDIPLCASASSFECCGPOVEPKCGRVGCGVAPNPHSMPQISLRFTGQHFQCGT 612
QY 228 DACSSKAVVSLRCIACG-VNLNSR-OSRIVGGSALPGAMPQVSLHYQ-NV-HVCGGS 283
Db 613 LIAPENVLTAAHCLKS-SRP-EF--YKYL-GAHEEYINGLDVQE-ISAKLILEPNNR 666
QY 284 IITPEWIVTAAHCVEKPLNPMWHTAFAGILROSFMFYAGAYVEKYISHPNYDSKTKNN 343
Db 667 DIALKLSRPATITDKYIPACLPSPNMYVADRTICYITGNETGTGAGR-LKEAQLPV 725
QY 344 DIALMKLQKPLTFNDLVKPVCLPMPGMLOPEQLCWISGWAIEKGTISEVLNAKAVL 403
Db 726 IENKVCN-RVEYLNNRYKSTELCAGOLAGVDSGQDGGPLYCFEKDYILQGVTSMGL 784
QY 404 IETORCNSRYVY-DNLITPAMICAGFLQGHVDSGQDGGPGLVTSKNINIMWLIGDTSMS 462
Db 785 GCAAPNKGYYVYVRSRFVDWIEREMNN 812
QY 463 GCAKAYRPGYGNVYFTDWIYROMRAD 490

Search completed: Wed Apr 19 23:45:02 2000
Job time : 24 secs.


 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution Rights by Oxford Molecular Ltd

MPERCH_PP protein - protein database search, using Smith-Waterman algorithm
 Run on: Wed Apr 19 23:22:52 2000; Maspar time 15.97 Seconds
 Tabular output not generated. 728.302 Million cell updates/sec

Title: >US-09-323-597-2
 Description: (1-491) from US09323597.pep
 Perfect Score: 3673
 Sequence: 1 MALMSGSPPAIGPYENHGY.....VYGNVMTFTDWIRQMRADG 491

Scoring table:
 PAM 150
 Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 08
 Listing first 45 summaries

Database: a-geneseq36
 1:geneseqp

Statistics: Mean 35.553; Variance 144.369; scale 0.246

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	895	24.4	798	1 R57283	Bovine enterokinase.	2.31e-77
2	710	19.3	233	1 W22966	Human serine protease	2.36e-58
3	697	19.0	418	1 R89435	Trypsin-like enzyme.	5.04e-57
4	691	18.8	356	1 W46917	Amino acid sequence of	2.07e-56
5	678	18.5	232	1 R89430	Trypsin-like enzyme.	4.38e-55
6	658	18.2	416	1 W96812	A mouse serine proteas	4.59e-54
7	653	17.8	285	1 W77301	Amino acid sequence of	1.55e-52
8	649	17.7	241	1 W22967	Human serine protease	3.95e-52
9	634	17.3	812	1 W07585	Murine plasminogen seq	1.33e-50
10	634	17.3	812	1 R83959	Complete mouse plasmin	1.33e-50
11	634	17.3	812	1 R83959	Murine plasminogen.	1.33e-50
12	615	16.7	289	1 W77302	Amino acid sequence of	1.13e-48
13	608	16.6	271	1 W77302	Amino acid sequence of	5.80e-48
14	608	16.6	317	1 Y13391	Amino acid sequence of	7.33e-48
15	607	16.5	270	1 W64239	Gerbil homolog of mo	7.33e-48
16	606	16.5	811	1 R12933	Plasminogen muten X1	9.26e-48
17	606	16.5	812	1 R12934	Plasminogen muten X2	9.26e-48
18	606	16.5	813	1 R12935	Plasminogen muten X3	9.26e-48
19	606	16.5	813	1 R12937	Plasminogen muten X6	9.26e-48
20	605	16.5	816	1 R12946	Plasminogen muten T19	1.17e-47
21	604	16.4	297	1 W77304	Amino acid sequence of	1.48e-47
22	603	16.4	810	1 R12938	Plasminogen muten T1	1.87e-47
23	603	16.4	811	1 R12943	Plasminogen muten T13	1.87e-47

24	603	16.4	811	1 R12945	Plasminogen muten T17	1.87e-47
25	603	16.4	811	1 R12944	Plasminogen muten T14	1.87e-47
26	603	16.4	811	1 R12939	Plasminogen muten T2	1.87e-47
27	603	16.4	812	1 R12948	Plasminogen muten X5	1.87e-47
28	604	16.4	813	1 R12936	Plasminogen muten X2	1.48e-47
29	597	16.3	437	1 W51457	Human plasminogen frag	7.57e-47
30	596	16.3	546	1 R34427	Sequence of tissue pla	5.98e-47
31	597	16.3	790	1 R60519	Human 'Glu' plasminoge	7.57e-47
32	597	16.3	791	1 W34285	Human plasminogen.	7.57e-47
33	597	16.3	810	1 W31169	Plasminogen protein fo	7.57e-47
34	597	16.3	810	1 R08065	Human plasminogen from	7.57e-47
35	597	16.3	810	1 R34428	Sequence encoded by a	7.57e-47
36	596	16.2	230	1 R56471	Mutant plasmin proteas	9.58e-47
37	594	16.2	314	1 W77296	Amino acid sequence of	1.52e-46
38	594	16.2	314	1 W97116	A human eosinophil ser	1.52e-46
39	594	16.2	314	1 Y13388	Amino acid sequence of	1.52e-46
40	595	16.2	816	1 R12942	Plasminogen muten T8	1.21e-46
41	593	16.1	230	1 R56470	Mutant plasmin proteas	1.92e-46
42	593	16.1	230	1 R49031	Plasmin protease domai	1.92e-46
43	593	16.1	243	1 R54766	Serine protease domain	1.92e-46
44	593	16.1	812	1 R12940	Plasminogen muten T6	1.92e-46
45	593	16.1	812	1 R12941	Plasminogen muten T7	1.92e-46

ALIGNMENTS

RESULT	ID	Accession	Standard	Protein	Length	AA
AC	R57283	standard	Protein	798	AA.	
DT	08-MAR-1995	(first entry)				
DE	Bovine enterokinase.					
KM	Enterokinase; EK; heavy chain; light chain; catalytic domain;					
KW	digestive disorder; cleavage; fusion protein; trypsinogen;					
KM	trypsin; enzyme; PACE gene.					
OS	Bos taurus.					
FH	Key					
FT	domain	Location/Qualifiers				
FT	domain	1..563				
FT	domain	/label="heavy-chain_C-terminal				
FT	domain	/note="non-catalytic domain"				
FT	domain	564..798				
FT	domain	/label="light-chain				
FT	domain	/note="catalytic domain"				
PD	MO9416083-A.					
PD	21-JUL-1994.					
PF	13-JAN-1994; U00616.					
PR	15-JAN-1993; US-005944.					
PA	(GEMT) GENETICS INST INC.					
PI	Levallic ER.					
DR	WPI: 94-249229/30.					
DR	N-PDB; 070104.					
PT	New nucleic acid encoding enterokinase activity - and related					
PT	vectors, host cells, expression products and antibodies are					
PT	useful in treating digestive disorders and for cleaving fusion					
PT	proteins					
PS	Disclosure: Page 28-30; 50pp; English.					
CC	The enterokinase (EK) (or the EK gene when used in gene therapy) is					
CC	used to treat digestive disorders associated with low EK activity					
CC	(esp. inability to process trypsinogen and for cleaving fusion					
CC	proteins, recombinant EK catalytic domain is much more					
CC	efficient than the native two-chain holoenzyme and is not					
CC	contaminated by other proteolytic enzymes. For expression of					
CC	recombinant EK, the 1691-2398 DNA fragment was fused to the 3'-end of					
CC	the signal peptide and pro-region of the human PACE gene. The prod.					
CC	could be expressed in CHO cells to produce a chimeric prod. from					
CC	which the pro-region as cleaved by endogenous PACE, providing mature					
CC	EK catalytic domain.					
CC	Sequence 798 AA:					
SO						
Query Match	24.4%	Score 895;	DB 1;	Length 798;		
Best Local Similarity	36.3%	Pred. No. 2.31e-77;				
Matches 138;	Conservative	84;	Mismatches 141;	Indels 17;	Gaps 16;	
DB	422 CKEDNFCQKD-GECIPVNLCDGFPHCKDSDAHCVLENGTDSGLVQFRIOSI-WR	479				


```

Oy 1201 gttgtctgcgccaacccagagatgatagtctctcaagccaaagaagctctgcctgattcttcggg 1260
Db 1228 gtgtgtctgcgccaacccagagatgatagtctctcaagccaaagaagctctgcctgattcttcggg 1287
Oy 1261 tgggggggcccacccgggagaaagggaagacactcagaaagtgtcttgaaacgctgcgccaagtgtctt 1320
Db 1288 tgggggggcccacccgggagaaagggaagacactcagaaagtgtcttgaaacgctgcgccaagtgtctt 1347
Oy 1311 ctcatctgaaacacagagatgtcaaacagagatatgtctatgtacaacactgtatcacacagcc 1380
Db 1348 ctcatctgaaacacagagatgtcaaacagagatatgtctatgtacaacactgtatcacacagcc 1407
Oy 1381 atgatactctgtccggctctccctcaggggaaagtcgattctcttcgcaagggttgacagtggagg 1440
Db 1408 atgatactctgtccggctctccctcaggggaaagtcgattctcttcgcaagggttgacagtggagg 1467
Oy 1441 cctctgtgtcactcttgaaagaacaatatctgtgtgtgctgaataggggatatacaagcttgggttct 1500
Db 1468 cctctgtgtcactcttgaaagaacaatatctgtgtgtgctgaataggggatatacaagcttgggttct 1527
Oy 1501 ggctgtgtccaaagctcttcacagacagagagtttaacgggaaatgtatgtatcttaacggagcttg 1560
Db 1528 ggctgtgtccaaagctcttcacagacagagagtttaacgggaaatgtatgtatcttaacggagcttg 1587
Oy 1561 atttatcgacaacatgaaagggcagagacgttatactccacatggtcttcgctcgttcgtgaagcttgctt 1620
Db 1588 atttatcgacaacatgaaagggcagagacgttatactccacatggtcttcgctcgttcgtgaagcttgctt 1647
Oy 1621 tacaaagaaacaaatctgggggtgtgttttctctcccgctgacatgattactcttaagagatgat 1680
Db 1648 tacaaagaaacaaatctgggggtgtgttttctctcccgctgacatgattactcttaagagatgat 1707
Oy 1681 tcagaaggtcactctattctttatataacagtgaaactgtctcggc 1723
Db 1708 tcagaaggtcactctattctttatataacagtgaaactgtctcggc 1750

RESULT 3
US-08-943-570-16
Sequence 16, Application US/08943570
GENERAL INFORMATION:
APPLICANT: COHEN, MAURICE
APPLICANT: COLPITTS, TRACEY L.
APPLICANT: FRIEDMAN, PAULA N.
APPLICANT: GRANADOS, EDWARD N.
APPLICANT: KLASS, MICHAEL R.
APPLICANT: RUSSELL, JOHN C.
APPLICANT: STEWART, KENT D.
APPLICANT: STROUPE, STEVEN D.
TITLE OF INVENTION: NOVEL SERINE PROMLEASE REAGENTS
TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
NUMBER OF SEQUENCES: 74
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories
STREET: 100 Abbott Park Road
CITY: Abbott Park
STATE: IL
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/943,570
FILING DATE:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:

```

[illegible]

RESULT 9
 ID W07585 standard: protein; 812 AA.
 AC W07585;
 DT 24-JUN-1997 (first entry)
 DE Murine plasminogen sequence.
 KW angiotatin; plasminogen; kringles; angiogenesis; cancer; arthritis;
 macular degeneration; diabetic retinopathy.
 OS Mus musculus.
 PN W09635774-A2.
 PD 14-NOV-1996.
 PR 26-APR-1996; U05856.
 PR 26-APR-1996; US-429743.
 PR 22-FEB-1996; US-605598.
 PR 08-MAR-1996; US-612788.
 PA (CHIL-) CHILDRENS MEDICAL CENT.
 PI Cao Y, Folkman MJ, Lin J, O'Reilly MS, Slin KL.
 DR WPI; 96-518662/51.
 PT Use of angiotatin fragments or aggregates - for inhibiting
 PT endothelial cell proliferation and treating angiogenesis-mediated
 PT diseases, e.g. cancer, arthritis or diabetic retinopathy
 PS Disclosure: Fig 1: 203pp; English.
 CC The invention relates to new methods and compositions for
 CC inhibiting endothelial cell proliferation, using as active component
 CC an angiotatin fragment, a combination of angiotatin fragments, or
 CC aggregate angiotatin. The fragment is preferably derived from murine,
 CC human, Rhesus, porcine or bovine plasminogen and is a kringles 1,
 CC kringles 2, kringles 3, kringles 2-3, kringles 1-3, kringles 1-2, kringles
 CC 1-4 or kringles 1-4BKUS protein. The aggregate angiotatin has a Mol. wt.
 CC of 45-65 kD and is derived from a plasminogen fragment beginning at
 CC approximately amino acid number 98 of murine, human, Rhesus, porcine or
 CC bovine plasminogen. The active component can be used for treating
 CC angiogenesis-mediated diseases such as cancer, arthritis, macular
 CC degeneration and diabetic retinopathy. It can also be used to develop
 CC antibodies for use in diagnosis, detection and therapy.
 CC The present sequence is the whole murine plasminogen sequence from
 CC which fragments or aggregates may be derived for use in the invention.
 SQ Sequence 812 AA;

Query Match 17.3%; Score 634; DB 1; Length 812;
 Best Local Similarity 41.4%; Pred. No. 1,33e-50; Indels 13; Gaps 12;
 Matches 111; Conservative 51; Mismatches 93;

DB 553 DYCDIPICASSAFECGKPEYEPKCGRVGCVANPHSPMPOISLRTFTGQHFCGCT 612
 QY 228 DASSKAVVSLRCLACG-VNLNSR-QSRIVGGSALPGAMPQVSLAHVQ-NV-HYCGGS 283
 DB 613 LIAPENVLTAAHCLKS-SRP-EF--YKVIL-GAHEEYIRGLDVOE-ISVAKLILEPNNR 666
 QY 284 IITPEVITVAHCEKELNPNMHTAFAGILROSFMEYGAQYVEKYSHPNDSKTKNN 343
 DB 667 DIALKLSRPATITDKYIPACLPSPNMTVADRTICITGNGEOTGFGAGR-LKEADLPV 725
 QY 344 DIALMKLQKPLTFNDIVKPVCLPFGMLOPEOLCWIISGNGATEKGTSEVLAAKVL 403
 DB 726 IEKVVCH-RVEYLNRRKSTELCAGLAGVDSQSGSPPLVCFEEDKYLIGVTSMGL 784
 QY 404 IETGRCHSRIVY-DNLITTPMICAGFLQGNVDSQSGSPPLVTSKNNIMLIDTISMG 462
 DB 785 GCARPNGVYVVRSEVDIEMERNNN 812
 QY 463 GCAKAYRPGYGVNMYVTDMYIRQMRAD 490

RESULT 10
 ID W94036 standard: protein; 812 AA.
 AC W94036;
 DT 12-APR-1999 (first entry)
 DE Murine plasminogen.
 KW Plasminogen; kringles; endothelial; angiogenesis; tumour; leukemia;
 rheumatoid arthritis; psoriasis; ocular angiogenic disease; ulcer;
 gene therapy; birth control; Crohn's disease; angiotatin.
 OS Mus sp.

FH Key Location/Qualifiers
 FT Domain 103..181
 FT /note="Kringles 1"
 FT Domain 185..262
 FT /note="Kringles 2"
 FT Domain 275..352
 FT /note="Kringles 3"
 FT Domain 377..454
 FT /note="Kringles 4"
 FT Domain 185..352
 FT /note="Kringles 2-3"
 FT Domain 103..352
 FT /note="Kringles 1-3"
 FT Domain 103..262
 FT /note="Kringles 1-2"
 FT Domain 103..454
 FT /note="Kringles 1-4"
 FT Domain 93..470
 FT /note="Kringles 1-4BKUS"

PN W09854217-A1.
 PD 03-DEC-1998.
 PE 29-MAY-1998; U10979.
 PR 30-MAY-1997; US-866735.
 PA (CHIL-) CHILDRENS MEDICAL CENT.
 PI Folkman MJ, O'Reilly MS;
 DR WPI; 99-059809/05.
 PT Use of plasminogen fragments - having an amino acid sequence similar
 PT to the kringles 1-5 region, for inhibiting endothelial cell
 PT proliferation and angiogenesis
 PS Examples: Fig 1A-B: 165pp; English.
 CC The invention relates to inhibition of endothelial cell proliferation
 CC that comprises administering to an endothelial cell a plasminogen
 CC fragment having an amino acid sequence similar to the kringles 1-5 region
 CC of a plasminogen molecule. The plasminogen fragments can be derived from
 CC murine, human, Rhesus, porcine or bovine plasminogens. The plasminogen
 CC fragments can be used for modulating angiogenesis and creating angiogenic
 CC mediated disease e.g. solid tumours, blood born tumours (leukaemia),
 CC tumour metastasis, benign tumours (e.g. haemangiomas, acoustic neuromas,
 CC neurofibromas, trachomas and pyogenic granulomas), rheumatoid arthritis,
 CC psoriasis, ocular angiogenic diseases (e.g. diabetic retinopathy,
 CC retinopathy of prematurity, macular degeneration, corneal graft
 CC rejection, neovascular glaucoma, retrolental fibroplasia, rubosis),
 CC Osler-Webber Syndrome, myocardial angiogenesis, haemophilic joints,
 CC plaque neovascularisation, telangiectasia, angiodioma, wound
 CC granulation. The fragments can also be used for treating excessive or
 CC abnormal stimulation of endothelial cells (intestinal adhesions), Crohn's
 CC disease, atherosclerosis, scleroderma and hypertrophic scars; as a birth
 CC control agent by preventing vascularisation required for embryo
 CC implantation; and in the treatment of diseases that have angiogenesis as
 CC a pathological consequence such as cat scratch disease (Rochelle minia
 CC guttosa) and ulcers (H. pylori). The nucleotide sequences encoding the
 CC plasminogen fragments can also be used for gene therapy. The products can
 CC be used for the production of antibodies and in detection and diagnosis.
 CC The present sequence represents a murine plasminogen sequence. The
 CC kringles regions that can be used in the invention are indicated in the
 CC features.
 SQ Sequence 812 AA;

Query Match 17.3%; Score 634; DB 1; Length 812;
 Best Local Similarity 41.4%; Pred. No. 1,33e-50; Indels 13; Gaps 12;
 Matches 111; Conservative 51; Mismatches 93;

DB 553 DYCDIPICASSAFECGKPEYEPKCGRVGCVANPHSPMPOISLRTFTGQHFCGCT 612
 QY 228 DASSKAVVSLRCLACG-VNLNSR-QSRIVGGSALPGAMPQVSLAHVQ-NV-HYCGGS 283
 DB 613 LIAPENVLTAAHCLKS-SRP-EF--YKVIL-GAHEEYIRGLDVOE-ISVAKLILEPNNR 666
 QY 284 IITPEVITVAHCEKELNPNMHTAFAGILROSFMEYGAQYVEKYSHPNDSKTKNN 343
 DB 667 DIALKLSRPATITDKYIPACLPSPNMTVADRTICITGNGEOTGFGAGR-LKEADLPV 725
 QY 344 DIALMKLQKPLTFNDIVKPVCLPFGMLOPEOLCWIISGNGATEKGTSEVLAAKVL 403


```
Db 624 CTGCAATGCGCGCTGACTCTATGGAACACAGATCAAGCCAGATGTCTGTGCTG 683
Qy 1398 cctgcaggaggaaacgtcgatctctgccagggtgcagtgagggccctcgtg----- 1449
Db 684 CCCGAGGGGTGGCATTGATGCCCTGCCAGGGCCGACAGCGGTGTCCTTTGTGTGAGGA 743
Qy 1449 ---cacttcgaagaacaatactcgtgcgtgatataggggatcacagctgaggctcgtg 1505
Db 744 CAGCATCTCTCGAGCCGACAGTTGGCGGCTGTGTGTCATTGTGAGTTGGGCACTGGCTG 803
Qy 1506 tgcacaagcttacagaccagagtgtagcggaaatgtagtattcacggaaatgattta 1565
Db 804 TGCCCTGGGCCAGAGCCAGAGCGCTTACACCAAGTCAGTGACTTCGGGAGTGAATCTT 863
Qy 1566 tcgacaatatgaggggcagagcgctaatcca 1594
Db 864 CCAGGCATTAAGACTCACTCCGAGCCA 892
```

Search completed: April 22, 2000, 06:03:00
Job time: 2649 sec